

Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) **EP 1 260 228 A2**

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
27.11.2002 Bulletin 2002/48

(51) Int Cl.⁷: **A61K 38/18**, A61K 48/00,
A61K 38/27

(21) Application number: 02011329.6

(22) Date of filing: 29.09.1997

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE
Designated Extension States:
AL LT LV RO SI

**(30) Priority: 01.10.1996 US 27633
16.01.1997 EP 97100583**

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
97944906.3 / 0 946 721

(71) Applicant: **Rappold-Hoerbrand, Gudrun**
69118 Heidelberg (DE)

(72) Inventors:

- **Rappold-Hoerbrand, Gudrun, Prof. Dr.**
69118 Heidelberg (DE)
- **Rao, Ercole, Dr.**
64560 Riedstadt (DE)

(74) Representative: Köster, Reinhold, Dr.
Werderplatz 9
69120 Heidelberg (DE)

Remarks:

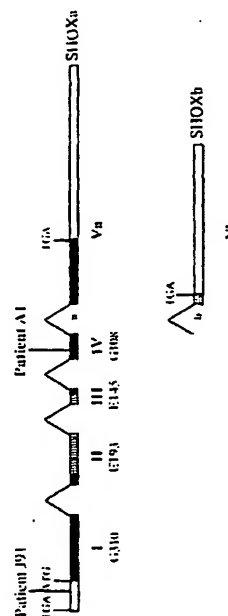
This application was filed on 23 - 05 - 2002 as a divisional application to the application mentioned under INID code 62.

(54) Human growth gene and short stature gene region. Therapeutic uses

(57) Subject of the present invention is an isolated human nucleic acid molecule encoding polypeptides containing a homeobox domain of sixty amino acids having the amino acid sequence of SEQ ID NO: 1 and having regulating activity on human growth.

Three novel genes residing within the about 500kb short stature critical region on the X and Y chromosome were identified. At least one of these genes is responsible for the short stature phenotype. The cDNA corresponding to this gene may be used in diagnostic tools, and to further characterize the molecular basis for the short stature-phenotype. In addition, the identification of the gene product of the gene provides new means and methods for the development of superior therapies for short stature.

Fig. 1



Description

[0001] The present invention relates to the isolation, identification and characterization of newly identified human genes responsible for disorders relating to human growth, especially for short stature or Turner syndrome, as well as the diagnosis and therapy of such disorders.

[0002] The isolated genomic DNA or fragments thereof can be used for pharmaceutical purposes or as diagnostic tools or reagents for identification or characterization of the genetic defect involved in such disorders. Subject of the present invention are further human growth proteins (transcription factors A, B and C) which are expressed after transcription of said DNA into RNA or mRNA and which can be used in the therapeutic treatment of disorders related to mutations in said genes. The invention further relates to appropriate cDNA sequences which can be used for the preparation of recombinant proteins suitable for the treatment of such disorders. Subject of the invention are further plasmid vectors for the expression of the DNA of these genes and appropriate cells containing such DNAs. It is a further subject of the present invention to provide means and methods for the genetic treatment of such disorders in the area of molecular medicine using an expression plasmid prepared by incorporating the DNA of this invention downstream from an expression promotor which effects expression in a mammalian host cell.

[0003] Growth is one of the fundamental aspects in the development of an organism, regulated by a highly organised and complex system. Height is a multifactorial trait, influenced by both environmental and genetic factors. Developmental malformations concerning body height are common phenomena among humans of all races. With an incidence of 3 in 100, growth retardation resulting in short stature account for the large majority of inborn deficiencies seen in humans.

[0004] With an incidence of 1:2500 life-born phenotypic females, Turner syndrome is a common chromosomal disorder (Rosenfeld et al., 1996). It has been estimated that 1-2% of all human conceptions are 45,X and that as many as 99 % of such fetuses do not come to term (Hall and Gilchrist, 1990; Robins, 1990). Significant clinical variability exists in the phenotype of persons with Turner syndrome (or Ullrich-Turner syndrome) (Ullrich, 1930; Turner, 1938). Short stature, however, is a consistent finding and together with gonadal dysgenesis considered as the lead symptoms of this disorder. Turner syndrome is a true multifactorial disorder. Both the embryonic lethality, the short stature, gonadal dysgenesis and the characteristic somatic features are thought to be due to monosomy of genes common to the X and Y chromosomes. The diploid dosis of those X-Y homologous genes are suggested to be requested for normal human development. Turner genes (or anti-Turner genes) are expected to be expressed in females from both the active and inactive X chromosomes or Y chromosome to ensure correct dosage of gene product. Haploinsufficiency (deficiency due to only one active copy), consequently would be the suggested genetic mechanism underlying the disease.

[0005] A variety of mechanisms underlying short stature have been elucidated so far. Growth hormone and growth hormone receptor deficiencies as well as skeletal disorders have been described as causes for the short stature phenotype (Martial et al., 1979; Phillips et al., 1981; Leung et al., 1987; Goddard et al., 1995). Recently, mutations in three human fibroblast growth factor receptor-encoding genes (FGFR 1-3) were identified as the cause of various skeletal disorders, including the most common form of dwarfism, achondroplasia (Shiang et al., 1994; Rousseau et al., 1994; Muenke and Schell, 1995). A well-known and frequent (1:2500 females) chromosomal disorder, Turner Syndrome (45,X), is also consistently associated with short stature. Taken together, however, all these different known causes account for only a small fraction of all short patients, leaving the vast majority of short stature cases unexplained to date.

[0006] The sex chromosomes X and Y are believed to harbor genes influencing height (Ogata and Matsuo, 1993). This could be deduced from genotype-phenotype correlations in patients with sex chromosome abnormalities. Cytogenetic studies have provided evidence that terminal deletions of the short arms of either the X or the Y chromosome consistently lead to short stature in the respective individuals (Zuffardi et al., 1982; Curry et al., 1984). More than 20 chromosomal rearrangements associated with terminal deletions of chromosome Xp and Yp have been reported that localize the gene(s) responsible for short stature to the pseudoautosomal region (PAR1) (Ballabio et al., 1989, Schaefer et al., 1993). This localisation has been narrowed down to the most distal 700 kb of DNA of the PAR1 region, with DXYS15 as the flanking marker (Ogata et al., 1992; 1995).

[0007] Mammalian growth regulation is organized as a complex system. It is conceivable that multiple growth promoting genes (proteins) interact with one another in a highly organized way. One of those genes controlling height has tentatively been mapped to the pseudoautosomal region PAR1 (Ballabio et al., 1989), a region known to be freely exchanged between the X and Y chromosomes (for a review see Rappold, 1993). The entire PAR1 region is approximately 2,700kb.

[0008] The critical region for short stature has been defined with deletion patients. Short stature is the consequence when an entire 700kb region is deleted or when a specific gene within this critical region is present in haploid state, is interrupted or mutated (as is the case with idiopathic short stature or Turner syndrome). The frequency of Turner's syndrome is 1 in 2500 females worldwide; the frequency of this kind of idiopathic short stature can be estimated to be 1 in 4.000 - 5.000 persons. Turner females and some short stature individuals usually receive an unspecific treatment with growth hormone (GH) for many years to over a decade although it is well known that they have normal GH levels

and GH deficiency is not the problem. The treatment of such patients is very expensive (estimated costs approximately 30.000 USD p.a.). Therefore, the problem existed to provide a method and means for distinguishing short stature patients on the one side who have a genetic defect in the respective gene and on the other side patients who do not have any genetic defect in this gene. Patients with a genetic defect in the respective gene - either a complete gene deletion (as in Turner syndrome) or a point mutation (as in idiopathic short stature) - should be susceptible for an alternative treatment without human GH, which now can be devised.

[0009] Genotype/phenotype correlations have supported the existence of a growth gene in the proximal part of Yp and in the distal part of Yp. Short stature is also consistently found in individuals with terminal deletions of Xp. Recently, an extensive search for male and female patients with partial monosomies of the pseudoautosomal region has been undertaken. On the basis of genotype-phenotype correlations, a minimal common region of deletion of 700 kb DNA adjacent to the telomere was determined (Ogata et al., 1992; Ogata et al., 1995). The region of interest was shown to lie between genetic markers DXYS20 (3cosPP) and DXYS15 (113D) and all candidate genes for growth control from within the PAR1 region (e.g., the hemopoietic growth factor receptor a; CSF2RA) (Gough et al., 1990) were excluded based on their physical location (Rappold et al., 1992). That is, the genes were within the 700 kb deletion region of the 2.700 kb PAR1 region.

[0010] Deletions of the pseudoautosomal region (PAR1) of the sex chromosomes were recently discovered in individuals with short stature and subsequently a minimal common deletion region of 700 kb within PAR1 was defined. Southern blot analysis on DNA of patients AK and SS using different pseudoautosomal markers has identified an Xp terminal deletion of about 700 kb distal to DXYS15 (113D) (Ogata et al, 1992; Ogata et al, 1995).

[0011] The gene region corresponding to short stature has been identified as a region of approximately 500 kb, preferably approximately 170 kb in the PAR1 region of the X and Y chromosomes. Three genes in this region have been identified as candidates for the short stature gene. These genes were designated SHOX (also referred to as SHOX93 or HOX93), (SHOX = short stature homeobox-containing gene), pET92 and SHOT (SHOX-like homeobox gene on chromosome three). The gene SHOX which has two separate splicing sites resulting in two variations (SHOX a and b) is of particular importance. In preliminary investigations, essential parts of the nucleotide sequence of the short stature gene could be analysed (SEQ ID No. 8). Respective exons or parts thereof could be predicted and identified (e.g. exon I [G310]; exon II [ET93]; exon IV [G108]; pET92). The obtained sequence information could then be used for designing appropriate primers or nucleotide probes which hybridize to parts of the SHOX gene or fragments thereof. By conventional methods, the SHOX gene can then be isolated. By further analysis of the DNA sequence of the genes responsible for short stature, the nucleotide sequence of exons I - V could be refined (v. fig. 1 - 3). The gene SHOX contains a homeobox sequence (SEQ ID NO: 1) of approximately 180 bp (v. fig. 2 and fig. 3), starting from the nucleotide coding for amino acid position 117 (Q) to the nucleotide coding for amino acid position 176 (E), i.e. from CAG (440) to GAG (619). The homeobox sequence is identified as the homeobox-pET93 (SHOX) sequence and two point mutations have been found in individuals with short stature in a German (A1) and a Japanese patient by screening up to date 250 individuals with idiopathic short stature. Both point mutations were found at the identical position and leading to a protein truncation at amino acid position 195, suggesting that there may exist a hot spot of mutation. Due to the fact that both mutations found, which lead to a protein truncation, are at the identical position, it is possible that a putative hot spot of recombination exists with exon 4 (G108). Exon specific primers can therefore be used as indicated below, e.g. GCA CAG CCA ACC ACC TAG (for) or TGG AAA GGC ATC ATC CGT AAG (rev).

[0012] The above-mentioned novel homeobox-containing gene, SHOX, which is located within the 170 kb interval, is alternatively spliced generating two proteins with diverse function. Mutation analysis and DNA sequencing were used to demonstrate that short stature can be caused by mutations in SHOX.

[0013] The identification and cloning of the short stature critical region according to the present invention was performed as follows: Extensive physical mapping studies on 15 individuals with partial monosomy in the pseudoautosomal region (PAR1) were performed. By correlating the height of those individuals with their deletion breakpoints a short stature (SS) critical region of approximately 700 kb was defined. This region was subsequently cloned as an overlapping cosmid contig using yeast artificial chromosomes (YACs) from PAR 1 (Ried et al., 1996) and by cosmid walking. To search for candidate genes for SS within this interval, a variety of techniques were applied to an approximately 600 kb region between the distal end of cosmid 56G10 and the proximal end of 51D11. Using cDNA selection, exon trapping, and CpG island cloning, the two novel genes were identified.

[0014] The position of the short stature critical interval could be refined to a smaller interval of 170 kb of DNA by characterizing three further specific individuals (GA, AT and RY), who were consistently short. To precisely localize the rearrangement breakpoints of those individuals, fluorescence *in situ* hybridization (FISH) on metaphase chromosomes was carried out using cosmids from the contig. Patient GA, with a terminal deletion and normal height, defined the distal boundary of the critical region (with the breakpoint on cosmid 110E3), and patient AT, with an X chromosome inversion and normal height, the proximal boundary (with the breakpoint on cosmid 34F5). The Y-chromosomal breakpoint of patient RY, with a terminal deletion and short stature, was also found to be contained on cosmid 34F5, suggesting that this region contains sequences predisposing to chromosome rearrangements.

[0015] The entire region, bounded by the Xp/Yp telomere, has been cloned as a set of overlapping cosmids. Fluorescence in situ hybridization (FISH) with cosmids from this region was used to study six patients with X chromosomal rearrangements, three with normal height and three with short stature. Genotype-phenotype correlations narrowed down the critical short stature interval to 270 kb of DNA or even less as 170 kb, containing the gene or genes with an important role in human growth. A minimal tiling path of six to eight cosmids bridging this interval is now available for interphase and metaphase FISH providing a valuable tool for diagnostic investigations on patients with idiopathic short stature.

Brief Description of the Drawings

[0016] Figure 1 is a gene map of the SHOX gene including five exons which are identified as follows: exon I: G310, exon II: ET93, exon III: ET45, exon IV: G108 and exons Va and Vb, whereby exons Va and Vb result from two different splicing sites of the SHOX gene. Exon II and III contain the homeobox sequence of 180 nucleotides.

[0017] Figures 2 and 3 are the nucleotide and predicted amino acid sequences of *SHOXa* and *SHOXb*:

SHOX a: The predicted start of translation begins at nucleotide 92 with the first in-frame stop codon (TGA) at nucleotides 968 - 970, yielding an open reading frame of 876 bp that encodes a predicted protein of 292 amino acids (designated as transcription factor A or *SHOXa* protein, respectively). An in-frame, 5' stop codon at nucleotide 4, the start codon and the predicted termination stop codon are in bold. The homeobox is boxed (starting from amino acid position 117 (Q) to 176 (E), i.e. CAG thru GAG in the nucleotide sequence). The locations of introns are indicated with arrows. Two putative polyadenylation signals in the 3'untranslated region are underlined.

SHOX b: An open reading frame of 876 bp exists from A in the first methionin at nucleotide 92 to the in-frame stop codon at nucleotide 767-769, yielding an open reading frame of 675 bp that encodes a predicted protein of 225 amino acids (transcription factor B or *SHOXb* protein, respectively). The locations of introns are indicated with arrows. Exons I-IV are identical with *SHOXa*, exon V is specific for *SHOX b*. A putative polyadenylation signal in the 3' untranslated region is underlined.

[0018] Figure 4 are the nucleotide and predicted amino acid sequence of SHOT. The predicted start of translation begins at nucleotide 43 with the first in-frame stop codon (TGA) at nucleotides 613 - 615, yielding an open reading frame of 573 bp that encodes a predicted protein of 190 amino acids (designated as transcription factor C or SHOT protein, respectively). The homeobox is boxed (starting from amino acid position 11 (Q) to 70 (E), i.e. CAG thru GAG in the nucleotide sequence). The locations of introns are indicated with arrows. Two putative polyadenylation signals in the 3'untranslated region are underlined.

[0019] Figure 5 gives the exon/intron organization of the human SHOX gene and the respective positions in the nucleotide sequence.

Brief Description of the SEQ ID:

[0020]

SEQ ID NO. 1: translated amino acid sequence of the homeobox domain (180 bp)

SEQ ID NO. 2: exon II (ET93) of the SHOX gene

SEQ ID NO. 3: exon I (G310) of the SHOX gene

SEQ ID NO. 4: exon III (ET45) of the SHOX gene

SEQ ID NO. 5: exon IV (G108) of the SHOX gene

SEQ ID NO. 6: exon Va of the SHOX gene

SEQ ID NO. 7: exon Vb of the SHOX gene

SEQ ID NO. 8: preliminary nucleotide sequence of the SHOX gene

SEQ ID NO. 9: ET92 gene

SEQ ID NO. 10: *SHOXa* sequence (see also fig. 2)

SEQ ID NO. 11: transcription factor A (see also fig. 2)

SEQ ID NO. 12: *SHOXb* sequence (see also fig. 3)

SEQ ID NO. 13: transcription factor B (see also fig 3)

SEQ ID NO. 14: SHOX gene

SEQ ID NO. 15: SHOT sequence (see also fig. 4)

SEQ ID NO. 16: transcription factor C (see also fig. 4)

[0021] Since the target gene leading to disorders in human growth (e.g. short stature region) was unknown prior to the present invention, the biological and clinical association of patients with this deletion could give insights to the function of this gene. In the present study, fluorescence in situ hybridization (FISH) was used to examine metaphase and interphase lymphocyte nuclei of six patients. The aim was to test all cosmids of the overlapping set for their utility as FISH probes and to determine the breakpoint regions in all four cases, thereby determining the minimal critical region for the short stature gene.

[0022] Duplication and deletion of genomic DNA can be technically assessed by carefully controlled quantitative PCR or dose estimation on Southern blots or by using RFLPs. However, a particularly reliable method for the accurate distinction between single and double dose of markers is FISH, the clinical application of is presently routine. Whereas in interphase FISH, the pure absence or presence of a molecular marker can be evaluated, FISH on metaphase chromosomes may provide a semi-quantitative measurement of inter-cosmid deletions. The present inventor has determined that deletions of about 10 kb (25% of signal reduction) can still be detected. This is of importance, as practically all disease genes on the human X chromosome have been associated with smaller and larger deletions in the range from a few kilobases to several megabases of DNA (Nelson et al., 1995).

[0023] Subject of the present invention are therefore DNA sequences or fragments thereof which are part of the genes responsible for human growth (or for short stature, respectively, in case of genetic defects in these genes). Three genes responsible for human growth were identified: SHOX, pET92 and SHOT. DNA sequences or fragments of these genes, as well as the respective full length DNA sequences of these genes can be transformed in an appropriate vector and transfected into cells. When such vectors are introduced into cells in an appropriate way as they are present in healthy humans, it is devisable to treat diseases involved with short stature, i.e. Turners syndrome, by modern means of gene therapy. For example, short stature can be treated by removing the respective mutated growth genes responsible for short stature. It is also possible to stimulate the respective genes which compensate the action of the genes responsible for short stature, i.e. by inserting DNA sequences before, after or within the growth/short stature genes in order to increase the expression of the healthy alleles. By such modifications of the genes, the growth/short stature genes become activated or silent, respectively. This can be accomplished by inserting DNA sequences at appropriate sites within or adjacent to the gene, so that these inserted DNA sequences interfere with the growth/short stature genes and thereby activate or prevent their transcription. It is also devisable to insert a regulatory element (e.g. a promotor sequence) before said growth genes to stimulate the genes to become active. It is further devisable to stimulate the respective promotor sequence in order to overexpress - in the case of Turner syndrome - the healthy functional allele and to compensate for the missing allele. The modification of genes can be generally achieved by inserting exogenous DNA sequences into the growth gene / short stature gene via homologous recombination.

[0024] The DNA sequences according to the present invention can also be used for transformation of said sequences into animals, such as mammals, via an appropriate vector system. These transgenic animals can then be used for *in vivo* investigations for screening or identifying pharmaceutical agents which are useful in the treatment of diseases involved with short stature. If the animals positively respond to the administration of a candidate compound or agent, such agent or compound or derivatives thereof would be devisable as pharmaceutical agents. By appropriate means, the DNA sequences of the present invention can also be used in genetic experiments aiming at finding methods in order to compensate for the loss of genes responsible for short stature (knock-out animals).

[0025] In a further object of this invention, the DNA sequences can also be used to be transformed into cells. These cells can be used for identifying pharmaceutical agents useful for the treatment of diseases involved with short stature, or for screening of such compounds or library of compounds. In an appropriate test system, variations in the phenotype or in the expression pattern of these cells can be determined, thereby allowing the identification of interesting candidate agents in the development of pharmaceutical drugs.

[0026] The DNA sequences of the present invention can also be used for the design of appropriate primers which hybridize with segments of the short stature genes or fragments thereof under stringent conditions. Appropriate primer sequences can be constructed which are useful in the diagnosis of people who have a genetic defect causing short stature. In this respect it is noteworthy that the two mutations found occur at the identical position, suggesting that a mutational hot spot exists.

[0027] In general, DNA sequences according to the present invention are understood to embrace also such DNA sequences which are degenerate to the specific sequences shown, based on the degeneracy of the genetic code, or which hybridize under stringent conditions with the specifically shown DNA sequences.

[0028] The present invention encompasses especially the following aspects:

- a) An isolated human nucleic acid molecule encoding polypeptides containing a homeobox domain of sixty amino acids having the amino acid sequence of SEQ ID NO: 1 and having regulating activity on human growth.
- b) An isolated DNA molecule comprising the nucleotide sequence essentially as indicated in fig. 2, fig. 3 or fig. 4, and especially as shown in SEQ ID NO: 10, SEQ ID NO: 12 or SEQ ID NO: 15.
- c) DNA molecules capable of hybridizing to the DNA molecules of item b).

d) DNA molecules of item c) above which are capable of hybridization with the DNA molecules of item 2. under a temperature of 60 - 70 °C and in the presence of a standard buffer solution.

e) DNA molecules comprising a nucleotide sequence having a homology of seventy percent or higher with the nucleotide sequence of SEQ ID NO: 10, SEQ ID NO: 12 or SEQ ID NO: 15 and encoding a polypeptide having regulating activity on human growth.

f) Human growth proteins having the amino acid sequence of SEQ ID NO: 11, 13 or 16 or a functional fragment thereof.

g) Antibodies obtained from immunization of animals with human growth proteins of item f) or antigenic variants thereof.

h) Pharmaceutical compositions comprising human growth proteins or functional fragments thereof for treating disorders caused by genetic mutations of the human growth gene.

i) A method of screening for a substance effective for the treatment of disorders mentioned above under item h) comprising detecting messenger RNA hybridizing to any of the DNA molecules described in a) - e) so as to measure any enhancement in the expression levels of the DNA molecule in response to treatment of the host cell with that substance.

j) An expression vector or plasmid containing any of the nucleic acid molecules described in a) - e) above which enables the DNA molecules to be expressed in mammalian cells.

k) A method for the determination of the gene or genes responsible for short stature in a biological sample of body tissues or body fluids.

[0029] In the method k) above, preferably nucleotide amplification techniques, e.g. PCR, are used for detecting specific nucleotide sequences known to persons skilled in the art, and described, for example, by Mullis et al. 1986, Cold Spring Harbor Symposium Quant. Biol. 51, 263-273, and Saiki et al., 1988, Science 239, 487-491, which are incorporated herein by reference. The short stature nucleotide sequences to be determined are mainly those represented by sequences SEQ ID No. 2 to SEQ ID No. 7.

[0030] In principle, all oligonucleotide primers and probes for amplifying and detecting a genetic defect responsible for diminished human growth in a biological sample are suitable for amplifying a target short stature associated sequence. Especially, suitable exon specific primer pairs according to the invention are provided by table 1. Subsequently, a suitable detection, e.g. a radioactive or non-radioactive label is carried out.

Table 1:

Exon	Sense primer	Antisense primer	Product (bp)	Ta (°C)
5'-I (G310)	SP 1	ASP 1	194	58
3'-I (G310)	SP 2	ASP 2	295	58
II (ET93)	SP 3	ASP 3	262	76/72/68
III (ET45)	SP 4	ASP 4	120	65
IV (G108)	SP 5	ASP 5	154	62
Va (SHOXa)	SP 6	ASP 6	265	61

explanation of the abbreviations for the primers:

SP 1 : ATTTCCAATGGAAAGGCGTAAATAAC

SP2 : ACGGCTTTTGTATCCAAGTCTTTTG

SP3 : GCCCTGTGCCCTCCGCTCCC

SP4 : GGCTCTTCACATCTCTCTCTGCTTC

SP5 : CCACACTGACACCTGCTCCCTTTG

SP6 : CCCGCAGGTCCAGGCTCAGCTG

ASP1 : CGCCTCCGCCGTTACCGTCCTTG

ASP 2 : CCCTGGAGCCGGCGCGCAAAG

ASP 3 : CCCC GCCCCCGCCCCCGG

ASP 4 : CTTCAAGTCCCCCAGTCCCG

ASP 5 : CTAGGGATCTTCAGAGGAAGAAAAAG

ASP 6 : GCTGCGCGGCGGGTCAGAGCCCCAG

[0031] Also, a single stranded RNA can be used as target. Methods for reversed transcribing RNA into cDNA are also well known and described in Sambrook et al., Molecular Cloning: A Laboratory Manual, New York, Cold Spring Harbor Laboratory 1989. Alternatively, preferred methods for reversed transcription utilize thermostable DNA polymerases having RT activity.

[0032] Further, the technique described before can be used for selecting those person from a group of persons being of short stature characterized by a genetic defect and which allows as a consequence a more specific medical treatment.

[0033] In another subject of the present invention, the transcription factors A, B and C can be used as pharmaceutical agents. These transcription factors initiate a still unknown cascade of biological effects on a molecular level involved with human growth. These proteins or functional fragments thereof have a mitogenic effect on various cells. Especially, they have an osteogenic effect. They can be used in the treatment of bone diseases, such as e.g. osteoporosis, and especially all those diseases involved with disturbance in the bone calcium regulation.

[0034] As used herein, the term "isolated" refers to the original derivation of the DNA molecule by cloning. It is to be understood however, that this term is not intended to be so limiting and, in fact, the present invention relates to both naturally occurring and synthetically prepared sequences, as will be understood by the skilled person in the art.

[0035] The DNA molecules of this invention may be used in forms of gene therapy involving the use of an expression

plasmid prepared by incorporating an appropriate DNA sequence of this invention downstream from an expression promoter that effects expression in a mammalian host cell. Suitable host cells are procaryotic or eucaryotic cells. Procaryotic host cells are, for example, *E. coli*, *Bacillus subtilis*, and the like. By transfecting host cells with replicons originating from species adaptable to the host, that is, plasmid vectors containing replication starting point and regulator sequences, these host cells can be transfected with the desired gene or cDNA. Such vectors are preferably those having a sequence that provides the transfected cells with a property (phenotype) by which they can be selected. For example, for *E. coli* hosts the strain *E. coli* K12 is typically used, and for the vector either pBR322 or pUC plasmids can be generally employed. Examples for suitable promoters for *E. coli* hosts are trp promoter, lac promoter or lpp promoter. If desired, secretion of the expression product through the cell membrane can be effected by connecting a DNA sequence coding for a signal peptide sequence at the 5' upstream side of the gene. Eucaryotic host cells include cells derived from vertebrates or yeast etc.. As a vertebrate host cell, COS cells can be used (Cell, 1981, 23: 175 - 182), or CHO cells. Preferably, promoters can be used which are positioned 5' upstream of the gene to be expressed and having RNA splicing positions, polyadenylation and transcription termination sequences.

[0036] The transcription factors A, B and C of the present invention can be used to treat disorders caused by mutations in the human growth genes and can be used as growth promoting agents. Due to the polymorphism known in the case of eukaryotic genes, one or more amino acids may be substituted. Also, one or more amino acids in the polypeptides can be deleted or inserted at one or more sites in the amino acid sequence of the polypeptides of SEQ ID NO: 11, 13 or 16. Such polypeptides are generally referred to equivalent polypeptides as long as the underlying biological activity of the unmodified polypeptide remains essentially unchanged.

[0037] The present invention is illustrated by the following examples.

Example 1

Patients

[0038] All six patients studied had de novo sex chromosome aberrations.

[0039] CC is a girl with a karyotype 45,X/46,X psu dic (X) (Xqter → Xp22.3::Xp22.3 → Xqter). At the last examination at 6 1/2 years of age, her height was 114 cm (25 - 50 the % percentile). Her mother's height was 155 cm, the father was not available for analysis. For details, see Henke et al., 1991.

[0040] GA is a girl with a karyotype 46,X der X (3pter → 3p23::Xp22.3 → Xqter). At the last examination at 17 years, normal stature (159 cm) was observed. Her mother's height is 160 cm and her father's height 182 cm. For details, see Kulharya et al, 1995.

[0041] SS is a girl with a karyotype 46,X rea (X) (Xqter → Xq26 :: Xp22.3 → Xq26:). At 11 years her height remained below the 3rd percentile growth curve for Japanese girls; her predicted adult height (148.5 cm) was below her target height (163 cm) and target range (155 to 191 cm). For details, see Ogata et al, 1992.

[0042] AK is a girl with a karyotype 46,X rea (X) (Xqter → Xp22.3::Xp22.3 → Xp21.3:). At 13 years her height remained below the 2nd percentile growth curve for Japanese girls; her predicted adult height (142.8 cm) was below her target height (155.5 cm) and target range (147.5 - 163.5 cm). For details, see Ogata et al, 1995.

[0043] RY: the karyotype of the ring Y patient is 46,X,r(Y)/46,Xdic r(Y)/45,X[95:3:2], as examined on 100 lymphocytes; at 16 years of age his final height was 148; the heights of his three brothers are all in the normal range with 170 cm (16 years, brother 1), 164 cm (14 years, brother 2) and 128 cm (9 years, brother 3), respectively. Growth retardation of this patient is so severe that it would also be compatible with an additional deletion of the GCY locus on Yq.

[0044] AT: boy with ataxia and inv(X); normal height of 116 cm at age 7, parents' heights are 156 cm and 190 cm, respectively.

Patients for mutation analysis:

[0045] 250 individuals with idiopathic short stature were tested for mutations in SHOXa. The patients were selected on the following criteria: height for chronological age was below the 3rd centile of national height standards, minus 2 standard deviations (SDS); no causative disease was known, in particular: normal weight (length) for gestational age, normal body proportions, no chronic organic disorder, normal food intake, no psychiatric disorder, no skeletal dysplasia disorder, no thyroid or growth hormone deficiency.

Family A:

[0046] Cases 1 and 2 are short statured children of a German non-consanguineous family. The boy (case 1) was born at the 38th week of gestation by cesarian section. Birth weight was 2660 g, birth length 47 cm. He developed normally except for subnormal growth. On examination at the age of 6.4 years, he was proportionate small (106.8 cm,

-2.6 SDS) and obese (22.7 kg), but otherwise normal. His bone age was not retarded (6 yrs) and bone dysplasia was excluded by X-ray analysis. IGF-I and IGFBP-3 levels as well as thyroid parameters in serum rendered GH or thyroid hormone deficiency unlikely. The girl (case 2) was born at term by cesarian section. Birth weight was 2920 g, birth length 47 cm. Her developmental milestones were normal, but by the age of 12 months poor growth was apparent (length: 67 cm, -3.0 SDS). At 4 years she was 89.6 cm of height (-3.6 SDS). No dysmorphic features or dysproportions were apparent. She was not obese (13 kg). Her bone age was 3.5 years and bone dysplasia was excluded. Hormone parameters were normal. It is interesting to note that both the girl and the boy grow on the 50 percentile growth curve for females with Turner syndrome. The mother is the smallest of the family and has a mild rhizomelic dysproportion (142.3 cm, -3.8 SDS). One of her two sisters (150 cm, -2.5 SDS) and the maternal grandmother (153 cm, -2.0 SDS) are all short without any dysproportion. One sister has normal stature (167 cm, +0.4 SDS). The father's height is 166 cm (-1.8 SDS) and the maternal grandfather's height is 165 cm (-1.9 SDS). The other patient was of Japanese origin and showed the identical mutation.

Example 2

Identification of the short stature gene

A. In situ hybridization

a) Florescence in situ hybridization (FISH)

[0047] Florescence in situ hybridization (FISH) using cosmids residing in the Xp/Yp pseudoautosomal region (PAR1) was carried out. FISH studies using cosmids 64/75cos (LLNLc110H032), E22cos (2e2), F1/14cos (110A7), M1/70cos (110E3), P99F2cos (43C11), P99cos (LLNLc110P2410), B6cosb (1CRFc104H0425), F20cos (34F5), F21cos (1CRFc104G0411), F3cos2 (9E3), F3cos1 (11E6), P117cos (29B11), P6cos1 (1CRFc104P0117), P6cos2 (LLNLc110E0625) and E4cos (15G7) was carried out according to published methods (Lichter and Cremer, 1992). In short, one microgram of the respective cosmid clone was labeled with biotin and hybridized to human metaphase chromosomes under conditions that suppress signals from repetitive DNA sequences. Detection of the hybridization signal was via FITC-conjugated avidin. Images of FITC were taken by using a cooled charge coupled device camera system (Photometrics, Tucson, AZ).

b) Physical mapping

[0048] Cosmids were derived from Lawrence Livermore National Laboratory X- and Y-chromosome libraries and the Imperial Cancer Research Fund London (now Max Planck Institute for Molecular Genetics Berlin) X chromosome library. Using cosmids distal to DXYS15, namely E4cos, P6cos2, P6cos1, P117cos and F3cos1 one can determine that two copies are still present of E4cos, P6cos2, P6cos1 and one copy of P117cos and F3cos1. Breakpoints of both patients AK and SS map on cosmid P6cos1, with a maximum physical distance of 10 kb from each other. It was concluded that the abnormal X chromosomes of AK and SS have deleted about 630 kb of DNA.

[0049] Further cosmids were derived from the ICRF X chromosome specific cosmid library (ICRFc104), the Lawrence Livermore X chromosome specific cosmid library (LLNLc110) and the Y chromosome specific library (LLCO3'M'), as well as from a self-made cosmid library covering the entire genome. Cosmids were identified by hybridisation with all known probes mapping to this region and by using entire YACs as probes. To verify overlaps, end probes from several cosmids were used in cases in which overlaps could not be proven using known probes.

c) Southern Blot Hybridisation

[0050] Southern blot analysis using different pseudoautosomal markers has provided evidence that the breakpoint on the X chromosome of patient CC resides between DXYS20 (3cosPP) and DXYS60 (U7A) (Henke et al, 1991). In order to confirm this finding and to refine the breakpoint location, cosmids 64/75cos, E22cos, F1/14cos, M1/70cos, F2cos, P99F2cos and P99cos were used as FISH probes. The breakpoint location on the abnormal X of patient CC between cosmids 64/75cos (one copy) and F1/14cos (two copies) on the E22PAC could be determined. Patient CC with normal stature consequently has lost approximately 260-290 kb of DNA.

[0051] Southern blot hybridisations were carried out at high stringency conditions in Church buffer (0.5 M NaPi pH 7.2, 7% SDS, 1mM EDTA) at 65°C and washed in 40 mM NaPi, 1% SDS at 65°C.

d) FISH Analysis

[0052] Biotinylated cosmid DNA (insert size 32 - 45 kb) or cosmid fragments (10 - 16 kb) were hybridised to metaphase chromosomes from stimulated lymphocytes of patients under conditions as described previously (Lichter and Cremer, 1992). The hybridised probe was detected via avidin-conjugated FITC.

e) PCR Amplification

[0053] All PCRs were performed in 50 µl volumes containing 100 pg-200 ng template, 20 pmol of each primer, 200 µM dNTP's (Pharmacia), 1.5 mM MgCl₂, 75 mM Tris/HCl pH9, 20mM (NH₄)₂SO₄ 0.01% (w/v) Tween20 and 2 U of Goldstar DNA Polymerase (Eurogentec). Thermal cycling was carried out in a Thermocycler GeneE (Techne).

f) Exon Amplification

[0054] Four cosmid pools consisting of each four to five clones from the cosmid contigs were used for exon amplification experiments. The cosmids in each cosmid pool were partially digested with Sau3A. Gel purified fractions in the size range of 4-10 kb were cloned in the BamHI digested pSPL3B vector (Burn et al, 1995) and used for the exon amplification experiments as previously described (Church et al., 1994).

g) Genomic Sequencing

[0055] Sonificated fragments of the two cosmids LLOYNCO3'M' 15D10 and LLOYNCO3'M'34F5 were subcloned separately into M13mp18 vectors. From each cosmid library at least 1000 plaques were picked, M13 DNA prepared and sequenced using dye-terminators, ThermoSequenase (Amersham) and universal M13-primer (MWG-BioTech). The gels were run on ABI-377 sequencers and data were assembled and edited with the GAP4 program (Staden).

[0056] Of all six patients, GA had the least well characterized chromosomal breakpoint. The most distal markers previously tested for their presence or absence on the X were DXS1060 and DXS996, which map approximately 6 Mb from the telomere (Nelson et al., 1995). Several cosmids containing different gene sequences from within PAR1 (MIC2, ANT3, CSF2RA, and XE7) were tested and all were present on the translocation chromosome. Cosmids from within the short stature critical region e.g., chromosome, thereby placing the translocation breakpoint on cosmid M1/70cos. A quantitative comparison of the signal intensities of M1/70cos between the normal and the rearranged X indicates that approximately 70% of this cosmid is deleted.

TABLE 2

	CC	GA	AK	SS
64/75cos	-	-		
E22cos	-	-		
F1/14cos	+	-		
M1/70cos	+	(+)		
F2cos	+	+		
P99F2cos	+	+		
P99cos	+	+		
B6cos		+		
F20cos				
F21cos				
F3cos2				
F3cos1			-	-
P117cos			-	-
P6cos1			+	+
P6cos2			+	+

TABLE 2 (continued)

	CC	GA	AK	SS
E4cos			+	+
Table 2: This table summarizes the FISH data for the 16 cosmid tested on four patients.				
[-] one copy; indicates that the respective cosmid was deleted on the rearranged X, but present on the normal X chromosome				
[+] two copies; indicates that the respective cosmid is present on the rearranged and on the normal X chromosome				
[(+)] breakpoint region; indicates that the breakpoint occurs within the cosmid as shown by FISH				

[0057] In summary, the molecular analysis on six patients with X chromosomal rearrangements using fluorescence-labeled cosmid probes and in situ hybridization indicates that the short stature critical region can be narrowed down to a 270 kb interval, bounded by the breakpoint of patient GA from its centromere distal side and by patients AK and SS on its centromere proximal side.

[0058] Genotype-phenotype correlations may be informative and have been chosen to delineate the short stature critical interval on the human X and Y chromosome. In the present study FISH analysis was used to study metaphase spreads and interphase nuclei of lymphocytes from patients carrying deletions and translocations on the X chromosome and breakpoints within Xp22.3. These breakpoints appear to be clustered in two of the four patients (AK and SS) presumably due to the presence of sequences predisposing to chromosome rearrangements. One additional patient Ring Y has been found with an interruption in the 270 kb critical region, thereby reducing the critical interval to a 170 kb region.

[0059] By correlating the height of all six individuals with their deletion breakpoint, an interval of 170 kb was mapped to within the pseudoautosomal region, presence or absence of which has a significant effect on stature. This interval is bounded by the X chromosomal breakpoint of patient GA at 340 kb from the telomere (Xp tel) distally and by the breakpoints of patients AT and RY at 510/520 kb Xp tel proximally. This assignment constitutes a considerable reduction of the critical interval to almost one fourth of its previous size (Ogata et al., 1992; Ogata et al., 1995). A small set of six to eight cosmids are now available for FISH experiments to test for the prevalence and significance of this genomic locus on a large series of patients with idiopathic short stature.

B. Identification of the Candidate Short Stature Gene

[0060] To search for transcription units within the smallest 170 kb critical region, exon trapping and cDNA selection on six cosmids (110E3, F2cos, 43C11, P2410, 15D10, 34F5) was carried out. Three different positive clones (ET93, ET45 and G108) were isolated by exon trapping, all of which mapped back to cosmid 34F5. Previous studies using cDNA selection protocols and an excess of 25 different cDNA libraries had proven unsuccessful, suggesting that genes in this interval are expressed at very low abundance.

[0061] To find out whether any gene in this interval was missed, the nucleotide sequence of about 140 kb from this region of the PAR1 was determined, using the random M13 method and dye terminator chemistry. The cosmids for sequence analysis were chosen to minimally overlap with each other and to collectively span the critical interval. DNA sequence analysis and subsequent protein prediction by the "X Grail" program, version 1.3c as well as by the exon-trapping program FEXHB were carried out and confirmed all 3 previously cloned exons. No protein-coding genes other than the previously isolated one could be detected.

C. Isolation of the Short Stature Candidate Gene SHOX

[0062] Assuming that all three exon clones ET93, ET45 and G108 are part of the same gene, they were used collectively as probes to screen 14 different cDNA libraries from 12 different fetal (lung, liver, brain 1 and 2) and adult tissues (ovary, placenta 1 and 2, fibroblast, skeletal muscle, bone marrow, brain, brain stem, hypothalamus, pituitary). Not a single clone among approximately 14 million plated clones was detected. To isolate the full-length transcript, 3' and 5'RACE were carried out. For 3'RACE, primers from exon G108 were used on RNA from placenta, skeletal muscle and bone marrow fibroblasts, tissues where G108 was shown to be expressed in. Two different 3'RACE clones of 1173 and 652 bp were derived from all three tissues, suggesting that two different 3'exons a and b exist. The two different forms were termed SHOXa and SHOXb.

[0063] To increase chances to isolate the complete 5'portion of a gene known to be expressed at low abundance, a HeLa cell line was treated with retinoic acid and phorbol ester PMA. RNA from such an induced cell line and RNA from

placenta and skeletal muscle were used for the construction of a 'Marathon cDNA library'. Identical 5'RACE cDNA clones were isolated from all three tissues.

Experimental procedure:

RT-PCR and cDNA Library Construction

[0064] Human polyA⁺RNA of heart, pancreas, placenta, skeletal muscle, fetal kidney and liver was purchased from Clontech. Total RNA was isolated from a bone marrow fibroblast cell line with TRIZOL reagent (Gibco-BRL) as described by the manufacturer. First strand cDNA synthesis was performed with the Superscript first strand cDNA synthesis kit (Gibco-BRL) starting with 100 ng polyA⁺RNA or 10µg total RNA using oligo(dt)-adapter primer (GGCCACGCGTC-GACTAGTAC[dT]₂₀N. After first strand cDNA synthesis the reaction mix was diluted 1/10. For further PCR experiments 5µl of this dilutions were used.

[0065] A 'Marathon cDNA library' was constructed from skeletal muscle and placenta polyA⁺RNA with the marathon cDNA amplification kit (Clontech) as described by the manufacturer.

[0066] Fetal brain (catalog # HL5015b), fetal lung (HL3022a), ovary (HL1098a), pituitary gland (HL1097v) and hypothalamus (HL1172b) cDNA libraries were purchased from Clontech. Brain, kidney, liver and lung cDNA libraries were part of the quick screen human cDNA library panel (Clontech). Fetal muscle cDNA library was obtained from the UK Human Genome Mapping Project Resource Center.

D. Sequence Analysis and Structure of SHOX Gene

[0067] A consensus sequence of SHOXa and SHOXb (1349 and 1870 bp) was assembled by analysis of sequences from the 5' and 3'RACE derived clones. A single open reading frame of 1870 bp (SHOXa) and 1349 bp (SHOXb) was identified, resulting in two proteins of 292 (SHOXa) and 225 amino acids (SHOXb). Both transcripts a and b share a common 5'end, but have a different last 3'exon, a finding suggestive of the use of alternative splicing signals. A complete alignment between the two cDNAs and the sequenced genomic DNA from cosmids LL0YNCO3"M"15D10 and LL0YNC3"M"34F5 was achieved, allowing establishment of the exon-intron structure (Fig.4). The gene is composed of 6 exons ranging in size from 58 bp (exon III) to 1146 bp (exon Va). Exon I contains a CpG-island, the start codon and the 5' region. A stop codon as well as the 3'-noncoding region is located in each of the alternatively spliced exons Va and Vb.

Example 3

[0068] Two cDNAs have been identified which map to the 160 kb region identified as critical for short stature. These cDNAs correspond to the genes SHOX and pET92. The cDNAs were identified by the hybridization of subclones of the cosmids to cDNA libraries.

[0069] Employing the set of cosmid clones with complete coverage of the critical region has now provided the genetic material to identify the causative gene. Positional cloning projects aimed at the isolation of the genes from this region are done by exon trapping and cDNA selection techniques. By virtue of their location within the pseudoautosomal region, these genes can be assumed to escape X-inactivation and to exert a dosage effect.

[0070] The cloning of the gene leading to short stature when absent (haploid) or deficient, represents a further step forward in diagnostic accuracy, providing the basis for mutational analysis within the gene by e.g. single strand conformation polymorphism (SSCP). In addition, cloning of this gene and its subsequent biochemical characterization has opened the way to a deeper understanding of biological processes involved in growth control.

[0071] The DNA sequences of the present invention provide a first molecular test to identify individuals with a specific genetic disorder within the complex heterogeneous group of patients with idiopathic short stature.

Example 4

Expression Pattern of SHOXa and SHOXb

[0072] Northern blot analysis using single exons as hybridisation probes revealed a different expression profile for every exon, strongly suggesting that the bands of different size and intensities represent cross-hybridisation products to other G,C rich gene sequences. To achieve a more realistic expression profile of both genes SHOXa and b, RT-PCR experiments on RNA from different tissues were carried out. Whereas expression of SHOXa was observed in skeletal muscle, placenta, pancreas, heart and bone marrow fibroblasts, expression of SHOXb was restricted to fetal kidney, skeletal muscle and bone marrow fibroblasts, with the far highest expression in bone marrow fibroblasts.

[0073] The expression of SHOXa in several cDNA libraries made of fetal brain, lung and muscle, of adult brain, lung and pituitary and of SHOXb in none of the tested libraries gives additional evidence that one spliced form (SHOXa) is more broadly expressed and the other (SHOXb) expressed in a predominantly tissue-specific manner.

[0074] To assess the transcriptional activity of SHOXa and SHOXb on the X and Y chromosome we used RT-PCR of RNA extracted from various cell lines containing the active X, the inactive X or the Y chromosome as the only human chromosomes. All cell lines revealed an amplification product of the expected length of 119 bp (SHOXa) and 541 bp (SHOXb), providing clear evidence that both SHOXa and b escape X-inactivation.

[0075] SHOXa and SHOXb encode novel homeodomain proteins. SHOX is highly conserved across species from mammalian to fish and flies. The very 5' end and the very 3' end - besides the homeodomain - are likely conserved regions between man and mouse, indicating a functional significance. Differences in those amino acid regions have not been allowed to accumulate during evolution between man and mouse.

Experimental procedures:

a) 5' and 3'RACE

[0076] To clone the 5' end of the SHOXa and b transcripts, 5'RACE was performed using the constructed 'Marathon cDNA libraries'. The following oligonucleotide primers were used: SHOX B rev, GAAAGGCATCCGTAAGGCTCCC (position 697-718, reverse strand [r]) and the adaptor primer AP1. PCR was carried out using touchdown parameters: 94°C for 2 min, 94°C for 30 sec, 70°C for 30 sec, 72°C for 2 min for 5 cycles. 94°C for 30 sec, 66°C for 30 sec, 72°C for 2 min for 5 cycles. 94°C for 30 sec, 62°C for 30 sec, 72°C for 2 min for 25 cycles. A second round of amplification was performed using 1/100 of the PCR product and the following nested oligonucleotide primers: SHOX A rev, GACGCCTTTATGCATCTGATTCTC (position 617-640 r) and the adaptor primer AP2. PCR was carried out for 35 cycles with an annealing temperature of 60°C.

[0077] To clone the 3' end of the SHOXa and b transcripts, 3'RACE was performed as previously described (Frohman et al., 1988) using oligo(dT)adaptor primed first strand cDNA. The following oligonucleotide primers were used: SHOX A for, GAATCAGATGCATAAAGGCGTC (position 619-640) and the oligo(dT)adaptor. PCR was carried out using following parameters: 94°C for 2 min, 94°C for 30 sec, 62°C for 30 sec, 72°C for 2 min for 35 cycles. A second round of amplification was performed using 1/100 of the PCR product and the following nested oligonucleotide primers: SHOX B for, GGGAGCCTTACGGATGCCTTTC (position 697-718) and the oligo(dT)adaptor. PCR was carried out for 35 cycles with annealing temperature of 62°C.

[0078] To validate the sequences of SHOXa and SHOXb transcripts, PCR was performed with a 5' oligonucleotide primer and a 3' oligonucleotide primer. For SHOXa the following primers were used: G310 for, AGCCCCGGCT-GCTCGCCAGC (position 59-78) and SHOX D rev, CTGCGCGGCGGGTCAGAGCCCCAG (position 959-982 r). For SHOXb the following primers were used: G310 for, AGCCCCGGCTGCTCGCCAGC and SHOX2A rev, GCCTCAG-CAGCAAAGCAAGATCCC (position 1215-1238 r). Both PCRs were carried out using touchdown parameters: 94°C for 2 min, 94°C for 30 sec, 70°C for 30 sec, 72°C for 2 min for 5 cycles. 94°C for 30 sec, 68°C for 30 sec, 72°C for 2 min for 5 cycles. 94°C for 30 sec, 65°C for 30 sec, 72°C for 2 min for 35 cycles. Products were gel-purified and cloned for sequencing analysis.

b) SSCP Analysis

[0079] SSCP analysis was performed on genomic amplified DNA from patients according to a previously described method (Orita et al., 1989). One to five µl of the PCR products were mixed with 5 µl of denaturation solution containing 95% Formamid and 10mM EDTA pH8 and denaturated at 95°C for 10 min. Samples were immediately chilled on ice and loaded on a 10% Polyacrylamide gel (Acrylamide:Bisacrylamide = 37.5:1 and 29:1; Multislotgel, TGGE base, Qiagen) containing 2% glycerol and 1xTBE. Gels were run at 15°C with 500V for 3 to 5 hours and silver stained as described in TGGE handbook (Qiagen, 1993).

c) Cloning and Sequencing of PCR Products

[0080] PCR products were cloned into pMOSBlue using the pMOSBlueT- Vector Kit from Amersham. Overnight cultures of single colonies were lysed in 100 µl H₂O by boiling for 10 min. The lysates were used as templates for PCRs with specific primers for the cloned PCR product. SSCP of PCR products allowed the identification of clones containing different alleles. The clones were sequenced with CY5 labelled vector primers Uni and T7 by the cycle sequencing method described by the manufacturer (ThermoSequenase Kit (Amersham)) on an ALF express automated sequencer (Pharmacia).

d) PCR Screening of cDNA Libraries

[0081] To detect expression of SHOXa and b, a PCR screening of several cDNA libraries and first strand cDNAs was carried out with SHOXa and b specific primers. For the cDNA libraries a DNA equivalent of 5×10^8 pfu was used. For SHOXa, primers SHOX E rev, GCTGAGCCTGGACCTGTTGGAAAGG (position 713-737 r) and SHOX a for were used. For SHOXb, the following primers were used: SHOX B for and SHOX2A rev. Both PCRs were carried out using touchdown parameters: 94°C for 2 min; 94°C for 30 sec, 68°C for 30 sec, 72°C for 40 sec for 5 cycles. 94°C for 30 sec, 65°C for 30 sec, 72°C for 40 sec for 5 cycles. 94°C for 30 sec, 62°C for 30 sec, 72°C for 40 sec for 35 cycles.

e) PCR Screening of cDNA Libraries

[0082] To detect expression of SHOXa and b, a PCR screening of several cDNA libraries and first strand cDNAs was carried out with SHOXa and b specific primers. For the cDNA libraries a DNA equivalent of 5×10^8 pfu was used. For SHOXa, primers SHOX E rev, GCTGAGCCTGGACCTGTTGGAAAGG (position 713-737 r) and SHOX a for were used. For SHOXb, the following primers were used: SHOX B for and SHOX2A rev. Both PCRs were carried out using touchdown parameters: 94°C for 2 min; 94°C for 30 sec, 68°C for 30 sec, 72°C for 40 sec for 5 cycles. 94°C for 30 sec, 65°C for 30 sec, 72°C for 40 sec for 5 cycles. 94°C for 30 sec, 62°C for 30 sec, 72°C for 40 sec for 35 cycles.

Example 5

Expression pattern of OG12, the putative mouse homolog of both SHOX and SHOT

[0083] In situ hybridisation on mouse embryos ranging from day 5 p.c. and day 18,5 p.c., as well as on fetal and newborn animals was carried out to establish the expression pattern. Expression was seen in the developing limb buds, in the mesoderm of nasal processes which contribute to the formation of the nose and palate, in the eyelid, in the aorta, in the developing female gonads, in the developing spinal cord (restricted to differentiating motor neurons) and brain. Based on this expression pattern and on the mapping position of its human homolog SHOT, SHOT represents a likely candidate for the Cornelia de Lange syndrome which includes short stature.

Example 6

[0084] Isolation of a novel SHOX-like homeobox gene on chromosome three, SHOT, being related to human growth / short stature

[0085] A new gene called SHOT (for SHOX-homolog on chromosome three) was isolated in human, sharing the most homology with the murine OG12 gene and the human SHOX gene. The human SHOT gene and the murine OG12 genes are highly homologous, with 99 % identity at the protein level. Although not yet proven, due to the striking homology between SHOT and SHOX (identity within the homeodomain only), it is likely that SHOT is also a gene likely involved in short stature or human growth.

[0086] SHOT was isolated using primers from two new human ESTs (HS 1224703 and HS 126759) from the EMBL database, to amplify a reverse-transcribed RNA from a bone marrow fibroblast line (Rao et al, 1997). The 5' and 3' ends of SHOT were generated by RACE-PCR from a bone marrow fibroblast library that was constructed according to Rao et al., 1997. SHOT was mapped by FISH analysis to chromosome 3q25/q26 and the murine homolog to the syntenic region on mouse chromosome 3. Based on the expression pattern of OG12, its mouse homolog, SHOT represents a candidate for the Cornelia Lange syndrome (which shows short stature and other features, including cranio-facial abnormalities) mapped to this chromosomal interval on 3q25/26.

Example 7

Searching for Mutations in Patients with Idiopathic Short Stature

[0087] The DNA sequences of the present invention are used in PCR, LCR, and other known technologies to determine if such individuals with short stature have small deletions or point mutations in the short stature gene.

[0088] A total of initially 91 (in total 250 individuals) unrelated male and female patients with idiopathic short stature (idiopathic short stature has an estimated incidence of 2 - 2,5 % in the general population) were tested for small rearrangements or point mutations in the SHOXa gene. Six sets of PCR primers were designed not only to amplify single exons but also sequences flanking the exon and a small part of the 5'UTR. For the largest exon, exon one, two additional internal-exon primers were generated. Primers used for PCR are shown in table 2.

[0089] Single strand conformation polymorphism (SSCP) of all amplified exons ranging from 120 to 295 bp in size

was carried out. Band mobility shifts were identified in only 2 individuals with short stature (Y91 and A1). Fragments that gave altered SSCP patterns (unique SSCP conformers) were cloned and sequenced. To avoid PCR and sequencing artifacts, sequencing was performed on two strands using two independent PCR reactions. The mutation in patient Y91 resides 28bp 5' of the start codon in the 5'UTR and involves a cytidine-to-guanine substitution. To find out if this mutation represents a rare polymorphism or is responsible for the phenotype by regulating gene expression e.g. through a weaker binding of translation initiation factors, his parents and a sister were tested. As both the sister and father with normal height also show the same SSCP variant (data not shown), this base substitution represents a rare polymorphism unrelated to the phenotype.

[0090] Cloning and sequencing of a unique SSCP conformer for patient A1 revealed a cytidine-to-thymidine base transition (nucleotide 674) which introduces a termination codon at amino-acid position 195 of the predicted 225 and 292 amino-acid sequences, respectively. To determine whether this nonsense mutation is genetically associated with the short stature in the family, pedigree analysis was carried out. It was found that all six short individuals (defined as height below 2 standard deviations) showed an aberrant SSCP shift and the cytidine-to-thymidine transition. Neither the father, nor one aunt and maternal grandfather with normal height showed this mutation, indicating that the grandmother has transferred the mutated allele onto two of her daughters and her two grandchildren. Thus, there is concordance between the presence of the mutant allele and the short stature phenotype in this family.

[0091] The identical situation as indicated above was found in another short stature patient of Japanese origin.

Example 8

[0092] The DNA sequences of the present invention are used to characterize the function of the gene or genes. The DNA sequences can be used as search queries for data base searching of nucleic acid or amino acid databases to identify related genes or gene products. The partial amino acid sequence of SHOX93 has been used as a search query of amino acid databases. The search showed very high homology to many known homeobox proteins. The cDNA sequences of the present invention can be used to recombinantly produce the peptide. Various expression systems known to those skilled in the art can be used for recombinant protein production.

[0093] By conventional peptide synthesis (protein synthesis according to the Merrifield method), a peptide having the sequence CSKSFQKSKDGNGG was synthesized and polyclonal antibodies were derived in both rabbits and chicken according to standard protocols.

References

[0094] The following references are herein incorporated by reference.

[0095] Ashworth A, Rastan S, Lovell-Badge R, Kay G (1991): X-chromosome inactivation may explain the difference in viability of XO humans and mice. *Nature* 351: 406-408.

[0096] Ballabio A, Bardoni A, Carrozzo R, Andria G, Bick D, Campbell L, Hamel B, Ferguson-Smith MA, Gimelli G, Fraccaro M, Maraschio P, Zuffardi O, Guilo S, Camerino G (1989): Contiguous gene syndromes due to deletions in the distal short arm of the human X chromosome. *Proc Natl Acad Sci USA* 86:10001-10005.

[0097] Blagowidow N, Page DC, Huff D, Mennuti MT (1989): Ullrich-Turner syndrome in an XY female fetus with deletion of the sex-determining portion of the Y chromosome. *Am. J. med. Genet.* 34: 159-162.

[0098] Cantrell MA, Bicknell JN, Pagon RA et al. (1989): Molecular analysis of 46,XY females and regional assignment of a new Y-chromosome-specific probe. *Hum. Genet.* 83: 88-92.

[0099] Connor JM, Loughlin SAR (1989): Molecular genetics of Turner's syndrome. *Acta Paediatr. Scand. (Suppl.)* 356: 77-80.

[0100] Distèche CM, Casanova M, Saal H, Friedman C, Sybert V, Graham J, Thuline H, Page DC, Fellous M (1986): Small deletions of the short arm of the Y-chromosome in 46,XY females. *Proc Natl Acad Sci USA* 83:7841-7844.

[0101] Ferguson-Smith MA (1965): Karyotype-phenotype correlations in gonadal dysgenesis and their bearing on the pathogenesis of malformations. *J. med. Genet.* 2: 142-155.

[0102] Ferrari D, Kosher RA, Dealy CN (1994): Limb mesenchymal cells inhibited from undergoing cartilage differentiation by a tumor promoting phorbol ester maintain expression of the homeobox-containing gene *MSX1* and fail to exhibit gap junctional communication. *Biochemical and Biophysical Research Communications.* 205(1): 429-434.

[0103] Fischer M, Bur-Romero P, Brown LG et al. (1990): Homologous ribosomal protein genes in the human X- and Y-chromosomes escape from X-inactivation and possible implementation for Turner syndrome. *Cell* 63: 1205-1218.

[0104] Freund C, Horsford DJ, McInnes RR (1996): Transcription factor genes and the developing eye: a genetic perspective. *Hum Mol Genet* 5: 1471-1488.

[0105] Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier A F, Resendez-Perez D, Affolter M, Otting G, Wüthrich K (1994): Homeodomain-DNA recognition. *Cell* 78: 211-223.

[0106] Gough NM, Gearing DP, Nicola NA, Baker E, Pritchard M, Callen DF, Sutherland GR (1990). Localization of

the human GM-CSF receptor gene to the X-Y pseudoautosomal region. *Nature* 345: 734736

[0107] Grumbach MM, Conte FA (1992): Disorders of sexual differentiation. In: Williams textbook of endocrinology, 8th edn., edited by Wilson JD, Foster DW, pp. 853-952, Philadelphia, WB Saunders.

[0108] Hall JG, Gilchrist DM (1990): Turner syndrome and its variants. *Pediatr. Clin. North Am.* 37: 1421-1436.

[0109] Henke A, Wapenaar M, van Ommen G-J, Maraschio P, Camerino O, Rappold GA (1991): Deletions within the pseudoautosomal region help map three new markers and indicate a possible role of this region in linear growth. *Am J Hum Genet* 49:811-819.

[0110] Hernandez D, Fisher EMC (1996): Down syndrome genetics: unravelling a multifactorial disorder. *Hum Mol Genet* 5:1411-1416.

[0111] Kenyon C (1994): If birds can fly, why can't we? Homeotic genes and evolution. *Cell* 78:175-180.

[0112] Krumlauf R (1994): *Hox* genes in vertebrate development. *Cell* 78: 191-201.

[0113] Kulharya AS, Roop H, Kukulich MK, Nachtman RG, Belmont JW, Garcia-Heras J (1995): Mild phenotypic effects of a de novo deletion Xpter → Xp22.3 and duplication 3pter → 3p23. *Am J Med Genet* 56:16-21.

[0114] Lawrence PA, Morata G (1994): Homeobox genes: their function in *Drosophila* segmentation and pattern formation. *Cell* 78: 181-189.

[0115] Lehrach H, Drmnac R, Hoheisel JD, Larin Z, Lemon G, Monaco AP, Nizetic D, et al., Hybridization finger printing in genome mapping and sequencing. In Davies KE, Tilghman S, Eds. *Genome Analysis 1990*: 39-81 Cold Spring Harbor, NY

[0116] Levilliers J, Quack B, Weissenbach J, Petit C (1989): Exchange of terminal portions of X- and Y-chromosomal short arms in human XY females. *Proc Natl Acad Sci USA* 86:2296-2300.

[0117] Lichter P, Cremer T, Human Cytogenetics: A practical Approach, IRL Press 1992, Oxford, New York, Tokyo

[0118] Lippe BM (1991): Turner Syndrome. *Endocrinol Metab Clin North Am* 20: 121-152. Magenis RE, Tochen ML Holahan KP, Carey T, Allen L, Brown MG (1984): Turner syndrome resulting from partial deletion of Y-chromosome short arm: localization of male determinants. *J Pediatr* 105: 916-919.

[0119] Nelson DL, Ballabio A, Cremers F, Monaco AP, Schlessinger D (1995).- Report of the sixth international workshop on the X chromosome mapping. *Cytogenet. Cell Genet.* 71: 308-342

[0120] Ogata T, Goodfellow P, Petit C, Aya M, Matsuo N (1992): Short stature in a girl with a terminal Xp deletion distal to DXYS15: localization of a growth gene(s) in the pseudoautosomal region. *J Med Genet* 29:455-459.

[0121] Ogata T, Tyler-Smith C, Purvis-Smith S, Turner G (1993): Chromosomal localisation of a gene(s) for Turner stigmata on Yp. *J. Med. Genet.* 30: 918-922.

[0122] Ogata T, Yoshizawa A, Muroya K, Matsuo N, Fukushima Y, Rappold GA, Yokoya S (1995): Short stature in a girl with partial monosomy of the pseudoautosomal region distal to DXYS15: further evidence for the assignment of the critical region for a pseudoautosomal growth gene(s). *J Med Genet* 32:831-834.

[0123] Ogata T, Matsuo N (1995): Turner syndrome and female sex chromosome aberrations: deduction of the principle factors involved in the development of clinical features. *Hum. Genet.* 95: 607-629.

[0124] Orita M, Suzuki Y, Sekiya T and Hayashi K (1989): Rapid and sensitive detection of point mutations and polymorphisms using the polymerase chain reaction. *Genomics* 5:874-879.

[0125] Pohlschmidt M, Rappold GA, Krause M, Ahlert D, Hosenfeld D, Weissenbach J, Gal A (1991): Ring Y chromosome: Molecular characterization by DNA probes. *Cytogenet Cell Genet* 56:65-68.

[0126] Qiagen (1993) TGGE Handbook, Diagen GmbH, TGMA 4112 3/93.

[0127] Rao E, Weiss B, Mertz A et al. (1995): Construction of a cosmid contig spanning the short stature candidate region in the pseudoautosomal region PAR 1. in: Turner syndrome in a life span perspective: Research and clinical aspects. Proceedings of the 4th International Symposium on Turner Syndrome, Gothenburg, Sweden, 18 - 21 May, 1995., edited by Albertsson-Wikland K, Ranke MB, pp. 19 - 24, Elsevier.

[0128] Rao E, Weiss B, Fukami M, Rump A, Niesler B, Mertz A, Muroya K, Binder G, Kirsch S, Winkelmann M, Nordsiek G, Heinrich U, Breuning MH, Ranke MB, Rosenthal A, Ogata T, Rappold GA (1997): Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nature Genet* 15:54-62

[0129] Rappold GA (1993): The pseudoautosomal region of the human sex chromosomes. *Hum Genet* 92:315-324.

[0130] Rappold GA, Willson TA, Henke A, Gough NM (1992): Arrangement and localization of the human GM-CSF receptor α chain gene CSF2RA within the X-Y pseudoautosomal region. *Genomics* 14:455-461.

[0131] Ried K, Mertz A, Nagaraja R, Trusnich M, Riley J, Anand R, Page D, Lehrach H., Elliso J, Rappold GA (1995): Characterization of a yeast artificial chromosome contig spanning the pseudoautosomal region. *Genomics* 29:787-792.

[0132] Robinson A (1990): Demography and prevalence of Turner syndrome. In: Turner Syndrome., edited by Rosenfeld RG, Grumbach MM, pp. 93 - 100, New York, Marcel Dekker.

[0133] Rosenfeld RG (1992): Turner syndrome: a guide for physicians. Second edition. The Turner's Syndrome Society.

[0134] Rosenfeld RG, Tesch L-G, Rodriguez-Rigau LJ, McCauley E, Albertsson-Wikland K, Asch R, Cara J, Conte

F, Hall JG, Lippe B, Nagel TC, Neely EK, Page DC, Ranke M, Saenger P, Watkins JM, Wilson DM (1994): Recommendations for diagnosis, treatment, and management of individuals with Turner syndrome. The Endocrinologist 4(5): 351-358.

[0135] Rovescalli AC, Asoh S, Nirenberg M (1996): Cloning and characterization of four murine homeobox genes. Proc Natl Acad Sci USA 93:10691-10696.

[0136] Schaefer L, Ferrero GB, Grillo A, Bassi MT, Roth EJ, Wapenaar MC, van Ommen G-JB, Mohandas TK, Rocchi M, Zoghbi HY, Ballabio A (1993): A high resolution deletion map of human chromosome Xp22. Nature genetics 4: 272-279.

[0137] Shalet SM (1993): Leukemia in children treated with growth hormone. Journal of Pediatric Endocrinology 6: 109-11

[0138] Vimpani GV, Vimpani AF, Lidgard GP, Cameron EHD, Farquhar JW (1977) Prevalence of severe growth hormone deficiency. Br Med J. 2: 427-430

[0139] Zinn AR, Page DC, Fisher EMC (1993): Turner syndrome: the case of the missing sex chromosome. TIG 9 (3): 90-93.

[0140] Preferred embodiments of the invention are especially the following:

E1. An isolated human nucleic acid molecule encoding polypeptides containing a homeobox domain of sixty amino acids having the amino acid sequence of SEQ ID NO: 1 and having regulating activity on human growth.

E2. A nucleic acid molecule according to embodiment E1 which is selected from the following group:

- a) an isolated DNA molecule comprising a nucleotide sequence (i) encoding a polypeptide containing a homeobox domain of sixty amino acids having the amino acid sequence of SEQ ID NO: 1 and which has the biological activity to regulate human growth, or (ii) encoding a polypeptide containing a homeobox domain of sixty amino acids having the amino acid sequence of SEQ ID NO: 1 except that one or more amino acid residues have been deleted, added or substituted but which retains the same biological activity of regulating human growth;
- b) an isolated DNA molecule comprising the nucleotide sequence of SHOX ET93 [SEQ ID NO: 2] and the nucleotide sequence of SHOX ET45 [SEQ ID NO: 4] or fragments thereof;
- c) nucleic acid molecules capable of hybridizing to the DNA molecules of a) or b); and
- d) DNA molecules comprising a nucleotide sequence having a homology of seventy percent or higher with the DNA molecules of a) or b).

E3. A DNA molecule according to embodiment E2 which encodes a polypeptide having an N-terminal and/or C-terminal amino acid extension to the homeobox domain of sixty amino acids of SEQ ID NO: 1.

E4. A DNA molecule according to embodiment E3 which encodes a polypeptide having a length of 150 to 350 amino acids.

E5. A DNA molecule according to any of embodiments E2 - E4 further comprising the nucleotide sequence of SHOX G310 [SEQ ID NO: 3].

E6. A DNA molecule according to any of embodiments 2 - 5 further comprising the nucleotide sequence of SHOX G108 [SEQ ID NO: 5].

E7. A DNA molecule according to any of embodiments E2 - E6 further comprising the nucleotide sequence of SHOX Va [SEQ ID NO: 6] or SHOX Vb [SEQ ID NO: 7].

E8. A DNA molecule according to any of embodiments E1 - E4 which encodes a polypeptide which is selected from the following group:

- a) transcription factor A having essentially the amino acid sequence of [SEQ ID NO: 11];
- b) transcription factor B having essentially the amino acid sequence of [SEQ ID NO: 13]; and
- c) transcription factor C having essentially the amino acid sequence of [SEQ ID NO: 16].

E9. DNA sequence comprising the nucleotide sequence of SHOX ET93 [SEQ ID No. 2].

E10. A DNA sequence according to embodiment E9 further comprising the nucleotide sequence of SHOX G310 [SEQ. ID NO. 3].

E11. A DNA sequence according to embodiments E9 or E10 further comprising the nucleotide sequence of SHOX ET45 [SEQ ID NO. 4].

E12. A DNA sequence according to any of embodiments E9 - E12 further comprising the nucleotide sequence of SHOX G108 [SEQ ID 5].

E13. A DNA sequence according to any of embodiments E9 - E12 further comprising either the nucleotide sequence of SHOX Va [SEQ ID 6] or SHOX Vb [SEQ ID 7].

E14. A DNA sequence according to embodiment E9 comprising the nucleotide sequence of SHOX ET93 [SEQ ID

No. 2] and the nucleotide sequence of SHOX ET45 [SEQ. ID. No. 4].

E15. A DNA sequence according to embodiment E9 comprising the nucleotide sequence of SHOX ET93 [SEQ ID NO 2], the nucleotide sequence of SHOX ET45 [SEQ. ID. No. 4] and the nucleotide sequence of SHOX G108 [SEQ ID 5].

E16. A DNA sequence according to any of embodiments E9 - E15 comprising the nucleotide sequences of SHOX G310 [SEQ ID NO. 3], SHOX ET93 [SEQ ID NO 2], SHOX ET45 [SEQ ID No. 4] and SHOX G108 [SEQ ID 5].

E17. A DNA sequence according to embodiment 17 further comprising the nucleotide sequence of SHOX Va [SEQ ID No 6].

E18. A DNA sequence according to embodiment E16 further comprising the nucleotide sequence of SHOX Vb [SEQ ID No. 7].

E19. A DNA sequence according to embodiment E9 consisting essentially of the isolated genomic sequence of the PAR1 region identified in [SEQ ID No. 14].

E20. A DNA sequence comprising the nucleotide sequence of SHOX ET92 [SEQ. ID No. 9].

E21. A DNA sequence according to any of embodiments E9 - E20 whereby the DNA is a genomic or isolated DNA responsible for regulating human growth.

E22. A DNA sequence according to any of embodiments E9 - E21 whereby the DNA is a cDNA.

E23. A cDNA according to embodiment E22 consisting essentially of the nucleotide sequence of SHOXa [SEQ ID No. 10] or SHOXb [SEQ ID NO. 12].

E24. A cDNA according to embodiment E22 consisting essentially of the nucleotide sequence of SHOT [SEQ ID No. 14].

E25. A human growth protein (transcription factor SHOXa) having the amino acid sequence given in [SEQ ID No. 11] or a functional fragment thereof.

E26. A human growth protein (transcription factor SHOXb) having the amino acid sequence given in [SEQ ID No. 13] or a functional fragment thereof.

E27. A human growth protein (transcription factor SHOT) having the amino acid sequence given in [SEQ ID NO: 16] or a functional fragment thereof.

E28. A cDNA encoding for a protein according to embodiment E25, E26 or E27.

E29. A pharmaceutical composition comprising a protein according to any of embodiments E25 to E27.

E30. A method for the treatment of short stature comprising administering to a subject in need thereof a therapeutically effective amount of a protein according to embodiment E25 to E27.

E31. Use of a protein according to embodiment E25 to E27 for the preparation of a pharmaceutical composition for the treatment of short stature.

E32. Use of a DNA sequence according to embodiments E1 - E24 for the preparation of a pharmaceutical composition for the treatment of disorders relating to mutations of the short stature gene.

E33. Use of a DNA sequence according to any of embodiments E1 - E24 for the preparation of a kit for the identification of individuals having a genetic defect responsible for diminished human growth.

E34. Use of a DNA sequence according to embodiment E33 for the identification of a gene responsible for short human stature.

E35. Method for the determination of short stature on the basis of RNA or DNA molecules, wherein the biological sample molecule to be examined is amplified in the presence of two nucleotide probes completely or in part complementary to any of the DNA sequences mentioned in SEQ ID No. 2 to SEQ ID No. 7 and subsequently determined by a suitable detection system.

E36. Use of the method according to embodiment E35 for the identification of persons having a genetic defect responsible for short stature.

E37. Transgenic animal transformed with a gene responsible for short stature containing a DNA sequence according to any one of embodiments E1 - E24.

E38. Cells transformed with a DNA sequence according to any one of embodiments E1 - E24.

E39. Test system for identifying or screening pharmaceutical agents useful for the treatment of human short stature comprising a cell according to embodiment E38.

E40. Method for identifying or screening of candidates for pharmaceutical agents useful for the treatment of disorders relating to mutations in the short stature gene comprising providing a test system according to embodiment E39 and determining variations in the phenotype of said cells or variations in the expression products of said cells after contacting said cells with said candidate pharmaceutical agents.

E41. An expression vector comprising a DNA molecule according to embodiments E1 - E8 which is capable of effecting the expression of the encoded polypeptide.

E42. A method for the *in vivo* treatment of human growth disorders related to at least one mutation in the SHOX or SHOT gene by gene therapy, comprising introducing into human cells an expression plasmid in which a DNA molecule according to any of embodiments E1 - E8 is incorporated downstream from the expression promoter that

EP 1 260 228 A2

effects expression in a human host cell.

E43. A method according to embodiment E42 for the treatment of Turner syndrome or short stature.

E44. Antibodies obtained by immunization of mammals using the transcription factors A, B or C or antigenic fragments thereof and isolating such antibodies from such mammals.

5

10

15

20

25

30

35

40

45

50

55

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Rappold-Hoerbrand, Gudrun, Dr.
 (B) STREET: Hausackerweg 14
 (C) CITY: Heidelberg
 (E) COUNTRY: Germany
 (F) POSTAL CODE (ZIP): 69118

(A) NAME: Rao, Ercole
 (B) STREET: Odenwaldstrasse 11
 (C) CITY: Riedstadt-Erfelden
 (E) COUNTRY: Germany
 (F) POSTAL CODE (ZIP): 64560

(ii) TITLE OF INVENTION: HUMAN GROWTH GENE AND SHORT STATURE GENE
 REGION

(iii) NUMBER OF SEQUENCES: 16

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/027,633
 (B) FILING DATE: 01-OCT-1996

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: EP 97100583.0
 (B) FILING DATE: 16-JAN-1997

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Gln	Arg	Arg	Ser	Arg	Thr	Asn	Phe	Thr	Leu	Glu	Gln	Leu	Asn	Glu	Leu
1				5					10					15	
Glu	Arg	Leu	Phe	Asp	Glu	Thr	His	Tyr	Pro	Asp	Ala	Phe	Met	Arg	Glu
			20					25					30		
Glu	Leu	Ser	Gln	Arg	Leu	Gly	Leu	Ser	Glu	Ala	Arg	Val	Gln	Val	Trp
		35				40						45			
Phe	Gln	Asn	Arg	Arg	Ala	Lys	Cys	Arg	Lys	Gln	Glu				
	50					55					60				

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 209 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "exon II: ET93"

(v) FRAGMENT TYPE: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GGATTTATGA ATGCAAAGAG AAGCGCGAGG ACGTGAAGTC GGAGGACGAG GACGGGCAGA	60
CCAAGCTGAA ACAGAGGCGC AGCCGCACCA ACTTCACGCT GGAGCAGCTG AACGAGCTCG	120
AGCGACTCTT CGACGAGACC CATTACCCCG ACGCCTTCAT GCGCGAGGAG CTCAGCCAGC	180
GCCTGGGGCT CTCCGAGGCG CGCGTGCAG	209

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 368 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "exon I: G310"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GTGATCCACC CGCGCGCAGG GCGCGTCTC TCCGCGCGGG GAGACGCGCG CATCCACCAG	60
CCCCGGCTGC TCGCCAGCCC CGGCCCCAGC CATGGAAGAG CTCACGGCTT TTGTATCCAA	120
GTCTTTTGAC CAGAAAAGCA AGGACGGTAA CGGCGGAGGC GGAGGCGGCG GAGGTAAGAA	180
GGATTCCATT ACGTACCGGG AAGTTTGTGA GAGCGGACTG GCGCGCTCCC GGGAGCTGGG	240
GACGTCGGAT TCCAGCCTCC AGGACATCAC GGAGGGCGGC GGCCACTGCC CGGTGCATTT	300
GTTCAAGGAC CACGTAGACA ATGACAAGGA GAACTGAAA GAATTCGGCA CCGCGAGAGT	360
GGCAGAAG	368

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "exon III: ET45"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GTTTGGTTCC AGAACCGGAG AGCCAAGTGC CGCAAACAAG AGAATCAGAT GCATAAAG 58

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "exon IV: G108"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GCGTCATCTT GGGCACAGCC AACCACCTAG ACGCCTGCCG AGTGGCACCC TACGTCAACA 60

TGGGAGCCTT ACGGATGCCT TTCCAACAG 89

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1166 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "exon : Va"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GTCCAGGCTC AGCTGCAGCT GGAAGGCGTG GCCCAGCGC ACCCGCACCT GCACCCGCAC 60

CTGGCGGCGC ACGCGCCCTA CCTGATGTTT CCCCCGCCG CCTTCGGGCT GCCCATCGCG 120

TCGTGGCCG AGTCCGCTC GGCCGCCGCC GTGGTCGCCG CCGCCGCCAA AAGCAACAGC 180

AAGAATTCCA GCATCGCCGA CCTGCGGCTC AAGGCGCGGA AGCACGCGGA GGCCCTGGGG 240

CTCTGACCCG CCGCGCAGCC CCCCAGCGCG CCGACTCCC GGGCTCCGCG CACCCCGCCT 300

GCACCGCGCG TCCTGCACTC AACCCCGCCT GGAGCTCCTT CCGCGGCCAC CGTGCTCCGG 360

GCACCCCGGG AGCTCCTGCA AGAGGCCTGA GGAGGGAGGC TCCCGGGACC GTCCACGCAC 420

GACCCAGCCA GACCCTCGCG GAGATGGTGC AGAAGGCGGA GCGGGTGAGC GGCCGTGCGT 480

CCAGCCCGGG CCTCTCCAAG GCTGCCCGTG CGTCTGGGA CCCTGGAGAA GGGTAAACCC 540

CCGCCTGGCT GCGTCTTCCT CTGCTATACC CTATGCATGC GGTAACTAC ACACGTTTGG 600

AAGATCCTTA GAGTCTATTG AAAGTCAAAA GATCCCGGAG CTGGTCTCCG ATGAAAATGC 660

CATTCTTCG TTGCCAACGA TTTCTTTAC TACCATGCTC CTTCCTTCAT CCCGAGAGGC 720

TGCGGAACGG GTGTGGATTT GAATGTGGAC TTCGGAATCC CAGGAGGCAG GGGCCGGGCT 780

CTCCTCCACC GCTCCCCCGG AGCCTCCAG GCAGCAATAA GGAAATAGTT CTCTGGCTGA 840

GGCTGAGGAC GTGAACCGCG GGCTTTGGAA AGGGAGGGGA GGGAGACCCG AACCTCCAC 900
 GTTGGGACTC CCACGTTCCG GGGACCTGAA TGAGGACCGA CTTTATAACT TTTCCAGTGT 960
 5 TTGATTCCCA AATGGGTCT GGTTTTGT TTGATTGGTA TTTTTTTTTT TTTTTTTTTT 1020
 TGCTGTGTTA CAGGATTCAG ACGCAAAGA CTGCATAAG AGACGGACGC GTGGTTGCAA 1080
 GGTGTCATAC TGATATGCAG CATTAAC TTT ACTGACATGG AGTGAAGTGC AATATTATAA 1140
 10 ATATTATAGA TTAATAAATA AATAGC 1166

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 625 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "exon Vb"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

25 ATGGAGTTT GCTCTGTGCG CCCAGGCTGG AGTATAATGG CATGATCTCG ACTCACTGCA 60
 ACCTCCGCT CCCGAGTTCA AGCGATTCTC CTGCCCTCAGC CTCCCAGTA GCTGGGATTA 120
 CAGGTGCCCA CCACCATGTC AAGATAATGT TTGTATTTTC AGTAGAGATG GGGTTTGACC 180
 30 ATGTTGGCCA GGCTGGTCTC GAACTCCTGA CCTCAGGTGA TCCACCCGCC TTAGCCTCCC 240
 AAAGTGCTGG GATGACAGGC GTGAGCCCT GCGCCCGGCC TTTGTAAC TTATTTTAAT 300
 TTTTTTTTTT TTTAAGAAA GACAGAGTCT TGCTCTGTCA CCCAGGCTGG AGCACACTGG 360
 35 TGCGATCATA GCTCACTGCA GCCTCAAACCT CCTGGGCTCA AGCAATCCTC CCACCTCAGC 420
 CTCTGAGTA GCTGGGACTA CAGGCACCCA CCACCACACC CAGCTAATTT TTTTGATTTT 480
 TACTAGAGAC GGGATCTTGC TTTGCTGCTG AGGCTGGTCT TGAGCTCCTG AGCTCCAAAG 540
 40 ATCCTCTCAC CTCCACCTCC CAAAGTGTTA GAATTACAAG CATGAACCAC TGCCCGTGGT 600
 CTCCAAAAAA AGGACTGTTA CGTGG 625

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15577 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "HOX93"

(ix) FEATURE:

- (A) NAME/KEY: exon
 (B) LOCATION: 1498..1807
 55 (D) OTHER INFORMATION: /function= "part of exon I (G310)"

- (ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:3844..4068
 (D) OTHER INFORMATION:/function= "pET92 region (first part)"
- (ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:4326..4437
 (D) OTHER INFORMATION:/function= "pET92 region (second part)"
- (ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:4545..4619
 (D) OTHER INFORMATION:/function= "pET92 region (third part)"
- (ix) FEATURE:
 (A) NAME/KEY: exon
 (B) LOCATION:5305..5512
 (D) OTHER INFORMATION:/function= "part of exon II (ET93)"
- (ix) FEATURE:
 (A) NAME/KEY: exon
 (B) LOCATION:11620..11729
 (D) OTHER INFORMATION:/function= "part of exon IV (G108)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CTCTCCCTGT TGTGTCTCTC TTTCTCTCTC TCCATCTCTC TCCGTCTTTC CCCCTCTGTC	60
TCTTTCTCTG TCTCCATCCC TCTGTCTCTC CCTTCTCTC TGTCTTTCCT TGTCTCTCTC	120
TTTCTCTCTC TCTCTCCATC TCTCTCTCTC CCGGTCTCTC TCTCTCCATC TCCCCGTCTC	180
TCCGTCTCTC TCTCTGCCCTC TCCCTGTCTG TCTCTCTCTT TGTGTGTGTT ACACACACCC	240
CAACCCACCG TCACTCATGT CCCCCACTG CTGTGCCATC TCACACAAGT TCACAGCTCA	300
GCTGTCTATC TGGGTCCCCA GGCCCCGCCG GGGAGGAAGA TGCGCCGTGG GGTACGGGA	360
GGAAGGGGAC TCCGGGCCCTC CTGGTGCCCC ACTTTATTTG CAGAAGGTCC TTGGCAGGAA	420
CCGTGACGCG TTTGGTTTCC AGGACTTGGA AAACGAATTT CAGGTCGCGA TGGCGAGCAC	480
CGGCTTCCCC TGAAGCACAT TCAATAGCGA GAGGCGGGAG GGAGCGAGCA GGAGCATCCC	540
ACCATGAAAA CCAAAAACAC AAGTATTTTT TTCACCCGGT AAATACCCCA GACGCCAGGG	600
TGACAGCGCG GCGCTAAGGG AGGAGGCCTC GCGCCGGGGT CCGCCGGGAT CTGGCGCGGG	660
CGGAAAGAAT ATAGATCTTT ACGAACCGBA TCTCCCGGGG ACCTGGGCTT CTTTCTGCGG	720
GCGCTGGAAG CCGGGGAGGC GGCCCCGGGG ATCCTCGGCC TCCGCCGCCG CCGCCTCCCA	780
AGCGCCCGCG TCCCGGTTTG GGGACACCCG GCCCCTTCTT CTCACTTTCG GGGATTCTCC	840
AGCCGCGTTC CATCTACCA ACTCTCCATC CAAGGGCGCG CCGCCACCAA CTGGAGCTC	900
ATCTTCTCCC AAAATCGTGC GTCCCCGGGG CGCCCCGGTC CCCCCCTCG CCATCTCAAC	960
CCCCGCGCGA CCGGGGCGCT TCCTGGAAAG ATCCAGGCGC CGGGCTCTGC GCTCCTCCCG	1020
GGAGCGAGGG CGGCCGACA ACTGGGACCC TCCTCTCTCC AGCCGTGAAC TCCTTGCTCTC	1080

EP 1 260 228 A2

	TCTGTCTCTC	TCTGCAGGAA	AACTGGAGTT	TGCTTTTCCT	CCGGCCACGG	AAAGAACGCG	1140
	GGTAACCTGT	GTGGGGGGCT	CGGGCGCCTG	CGCCCCCTC	CTGCGCGCGC	GCTCTCCCTT	1200
5	CCAAAAATGG	GATCTTTCCC	CCTTCGCACC	AAGGTGTACG	GACGCCAAAC	AGTGATGAAA	1260
	TGAGAAGAAA	GCCAATTGCC	GGCCTGGGGG	GTGGGGGAGA	CACAGCGTCT	CTGCGTGCCT	1320
	CCGCCGCGGA	GCCCGGAGAC	CAGTAATTGC	ACCAGACAGG	CAGCGCATGG	GGGGCTGGGC	1380
10	GAGGTGCGCG	CGTATAAATA	GTGAGATTTC	CAATGGAAG	GCGTAAATAA	CAGCGCTGGT	1440
	GATCCACCCG	CGCGCACGGG	CCGTCTCTCT	CGCGCGGGGA	GACGCGCGCA	TCCACCAGCC	1500
	CCGGCTGCTC	GCCAGCCCCG	GCCCCAGCCA	TGGAAGAGCT	CACGGCTTTT	GTATCCAAGT	1560
15	CTTTTGACCA	GAAAAGCAAG	GACGGTAACG	GCGGAGGCGG	AGGCGGCGGA	GGTAAGAAGG	1620
	ATTCCATTAC	GTACCGGGAA	GTTTTGGAGA	GCGGACTGGC	GCGCTCCCGG	GAGCTGGGGA	1680
	CGTCGGATTG	CAGCCTCCAG	GACATCACGG	AGGGCGGCGG	CCACTGCCCG	GTGCATTTGT	1740
20	TCAAGGACCA	CGTAGACAAT	GACAAGGAGA	AACTGAAAGA	ATTGCGCACC	GCGAGAGTGG	1800
	CAGAAAGTAA	GTTCTTTTGC	GCGCCGGCTC	CAGGGGGGCC	CTCCTGGGGT	TCGGCGCCTC	1860
	CTCGCCACGG	AGTCGGCCCC	GCGCGCCCTT	CGTGTTGCAC	ATTTGCAGCT	CCCGTCTCGC	1920
25	CAGGGTAAGG	CCCGGGCCGT	CAGGCTTTGC	CTAAGAAAGG	AAGGAAGGCA	GGAGTGGACC	1980
	CGACCGGAGA	CGCGGGTGGT	GGGTAGCGGG	GTGCGGGGGG	ACCCAGGGAG	GGTCGCAGCG	2040
	GGGGCCGCGC	GCGTGGGCAC	CGACACGGGA	AGGTCCCGGG	CTGGGGTGGG	TCCGGGTGGC	2100
30	TGTGCCTGAA	GCCGTAGGGC	CTGAGATGTC	TTTTTCATTT	TCTTTTCTTT	TCCTTTCTTT	2160
	TTTTTGTTTG	TTTGTTTGTT	TGTTTGAGAC	AGAGTCTCGC	TCTGTCCCCC	AGGCTGGAGT	2220
	GCAGTGGTGC	GATCTCGGCT	CACTGCAACC	TCCGCCTCCT	GGGTTCGAAG	GATTCTCCTG	2280
35	CCTCAGCCTC	CCCAGTAGCT	GGGATTACAG	GCATGCACCA	CCACGCCTGG	CTAATTTTGT	2340
	TGCTTTTAGT	AAAGACGGGG	ATTCAACCATG	TGCGCCAGGC	TGGTCTCGAA	CTCCTGACCT	2400
	CAGGTGATCC	ACCCGCCTCG	GCCTCCCAAA	GTGCTGGGAT	GACAGGCGTG	AGGCACCGCG	2460
40	CCCGGCCTGG	GTCTTGACGG	CTTAGGATGT	GTGTTTCTGT	CTCTGCCTGT	CTGCCTTGTA	2520
	TTTACGGTCA	CCCAGACGCA	CAGAGGAGCC	GTCTCCACGC	GCCTTCCCAG	CGCTCAGCGC	2580
	CTGCCGGGCC	CCCGGAGATC	ACGGGAAGAC	TCGAGGCTGC	GTGGTAGGAG	ACGGGAAGGC	2640
45	CCCGGGTCAG	CTCGGTTCCTG	TTTCNCTTTA	AGGAACCCCT	CATTATTATT	TCATTGTTTT	2700
	CCTTTGAACG	TCGAGGCTTG	ATCTTGCGGA	AAGCTGTTGG	GTCCATAAAA	ACCACTCCCG	2760
	TGAGCGGAGG	TGGCCGGGAT	CTGGATGGGG	CGCGAGGGGC	CCCGGGGAAG	CTGGCGGCTT	2820
50	CGCGGGCGCG	TCCTAAGTCA	AGGTTGTCAG	AGCGCAGCCG	GTTGTGCGCG	GCCCGGGGNN	2880
	AGCTCCCCCT	TGGCCCTTCC	TCCTGAGACC	TCAGTGGTGG	GTCGTCCCGT	GGTGGAATC	2940
	GGGGAGTAAG	AGGCTCAGAG	AGAGGGGCTG	GCCCCGGGGA	TCTCTGTGCA	CACACGACAA	3000
55	CTGGGCGGCA	TACATCTTAA	GAATAAAATG	GGCTGGCTGT	GTCGGGGCAC	AGCTGGAGAC	3060

EP 1 260 228 A2

	GGCTATGGAC	GCCTGTTATG	TTTTCATTAC	AAAGACGCAG	AGAATCTAGC	CTCGGCTTTT	3120
	GCTGATTCGC	AAAGTTGAGG	TGCGAGGGTG	AATGCCCCAA	AGGTAATTCT	TCCTAAGACT	3180
5	CTGGGGCTAC	CTGCTCTCCG	GGGCCCTGCA	TTTGGGGTGT	GGAGTGGCCC	CGGGAAATAG	3240
	CCCTTGTTAT	CGTAGGAGGC	ACCAGGCAGC	TTCCCAAGGC	CCTGACTTTG	TCGAAGCAGA	3300
	AAGCTGTGGC	TACGGTTTAC	AAAGCAGTCC	CCGGTTTCTG	ACCGTCTAAG	AGGCAGGAGC	3360
10	CCAGCCTGCC	TTTGACAGTG	AGAGGAGTTC	CTCCCTACAC	ACTGCTGCGG	GCACCCGGCA	3420
	CTGTAATTCA	TACACAGAGA	GTTGGCCTTC	CTGGACGCAA	GGCTGGGAGC	CGCTTGAGGG	3480
	CCTGCGTGTA	ATTTAAGAGG	GTTTCGCANGC	GCCCCGCGGC	CGCTTCTGNT	GGGGTTGCTT	3540
15	TTTGGTTGTC	CTTCNGCAAA	CACCGTTTTC	CTCCTCTNGN	AACCTCTCTT	TNCTCCCCCN	3600
	TGGCCNGTNG	GACCCGGGNA	NGAGCAAAGT	GTCTCCAGA	CCNTTTTGAA	ANGTGAGAGG	3660
	AAAATAAAGA	CCAGGCCAAA	NNGACCCAGG	GCCACAGGAG	AGGAGACAGA	GAGTCCCCGT	3720
20	TACATTTTNC	CCCTTGCGTG	GGTGACAGAA	GACCCCGGGG	CCAGGACTGC	CACCCAGGCT	3780
	ACTATTTATT	CATCAGATCC	AAGTTAAATC	GAGGTTGGAG	GGCAGGGGAG	AGTCTGAGGT	3840
	TACCGTGGAA	GCCTGGAGTT	TTTGGGNAAC	AGCGTGTCCC	CGCCGAGCCT	GGGAGCCCCG	3900
25	GGGTTCTGCA	AAGCCTGCGG	GTGTTTGAGG	ACTTTGAAGA	CCAGTTTGTC	AGTTGGGCTC	3960
	AATTNCCTGG	GGTTCAGACT	TAGAGAAATG	AAGGAGGGAG	AGCTGGGGTC	GTCTCCAGGA	4020
	AACGATTCAC	TTGGGGGGAA	GGAATGGAGT	GTTCTTGACG	GCACATGTCT	GTTAGGAGGT	4080
30	GAAACAGAAT	GTGAAATCCA	CGTTGGAGTA	AGCGTCCAGC	GCTGAATGTA	GCTCGGGGTG	4140
	GGGTGGGAGG	GCCCTGGTGT	GGATCGTGGA	AGGNAAGAAA	GACAGAACAG	GGTGCTAGTA	4200
	TTTACCCCGT	TNCCCTGTAG	ACACCCCTGA	TTTGTCAGCT	TTGCAAGCTT	CTTGGTTGCA	4260
35	GCGGCCTTGC	CTGTGCCCTT	TTGAGACTGT	TTCCAGACTA	AACCTCCAAA	TGTCAGCCCC	4320
	TTACCCCTGA	CAGCAAGGGA	CATCTCATTA	GGGCATCGCG	TGCTTCTCAT	CTGTGNCTCA	4380
	GCAGGCCCN	AGATAGGAAN	CANGAGGGGC	NGTTGGNAGA	TGCNCACTTC	CACCAGCCCT	4440
40	GGGNTTGAAG	GGGANGCGAN	GGGANGACNA	CCTTTTANCT	TAAACCCCTN	GAGCTTGGTN	4500
	CAGAGAGGNC	TGAATGTCTA	AAATGAGGAA	GAAAAGGTTT	TTACCTGGA	AACGCTTGAG	4560
	GGCTGAGTCT	TCTGCCCNIT	CTGACNTCCC	CCAGCAAATA	CAGACAGGTC	ACCAANTAC	4620
45	TGGAGATGAG	AAAGTGCCAT	TTTGGGCACA	CTCTGGTGGG	GTAGGTGCCC	GACCGCGTGT	4680
	GAAAAANGTG	GGAANNGGAG	AGATTTCTGN	CGCACGCGGT	TCAGCCCCCA	GGCGCGGNTG	4740
	GCNGCATTCN	AGGNTACTCA	GACGCGGTTC	TGCTGTTCTG	CTGAGAAACA	GGCTTCGGGT	4800
50	AGGGGCTCCT	AGCTCCGCCA	GATCGCGGAG	GGACCCCCAG	CCCTCCTGCG	CTGCAGCGGT	4860
	GGGGATAGCG	TCTCTCCGTA	GGCCTAGAAT	CTGCAACCCG	CCCCGGGTCC	TCCCCGTGTC	4920
	CTTCCCGGGC	GTCCCGCCGG	GGATCCCA	GTTGGCAGCT	CTTCTCAAA	TTCTTTCCCT	4980
55	TAAAAATAGG	ATTTGACACC	CCACTCTCCT	TAAAAA	AAATAAGAAA	AAAAGGTTAG	5040
	GTTATGTCAA	CAGAGGTGAA	GTGATAATT	GAGGAAACGA	TTCTGAGATG	AGGCCAAGAA	5100

EP 1 260 228 A2

	AACAACGCTC GTGCAAAGCC CAGGTTTTTG GGAAAGCAGC GAGTATCCTC CTCGGCTTTT	5160
	GCGTTATGGA CCCACGCAG TTTTTCGCTC AAAGCGCATT GGTTTTCGAG GGCCCCCTTT	5220
5	CCACCGCGGG ATGCACGAAG GGGTTCGCCA CGTTGCGCAA AACCTCCCCG GCCTCAGCCC	5280
	TGTGCCCTCC GCTCCCCACG CAGGGATTTA TGAATGCAAA GAGAAGCGCG AGGACGTGAA	5340
	GTCGAGGAC GAGGACGGG AGACCAAGCT GAAACAGAGG CGCAGCCGCA CCAACTTCAC	5400
10	GCTGGAGCAG CTGAACGAGC TCGAGCGACT TTTTGACGAG ACCCATTACC CCGACGCCTT	5460
	CATGCGCGAG GAGCTCAGCC AGCGCCTGGG GCTTTCGAG GCGCGCGTGC AGGTAGGAAC	5520
	CCGGGGGCGG GGGCGGGGG CCCGGAGCCA TCGCCTGGTC CTCGGGAGCG CACAGCACGC	5580
15	GTACAGCCAC CTGCGCCCGG GCCGCCCGCG TCCCTTCCC GGAGCGCGGG GAGGTGGGT	5640
	GAGGGACGGG CTGGGGTTCC TGGACTTTTG GAGACGCCTG AGGCCTGTAG GATGGGTTC	5700
	TTGCGTTTGT TTTTCACCAA CAGCAAACAA ATATATATAC ATATATATTA TACAAATAAC	5760
20	AAATAATAT ATATGTTATA CAGATGGGTA TATTGTATAT ATTATAGATA TTTGTTTCGTC	5820
	CTTGGTGCAA AGACACCCGG TGAACCCATA TATTGGCTCC TGAAGTCCCTT CGGTTCCCTT	5880
	GGGATTGGTT ATAGGGGCAA CACATGCAAA CAAACTTTC CCTGGATTAT ACTTAGGAGA	5940
25	CGAAGCTACA GATGCGTTTG ATCCAGAGTG TTTTACAAGA TTTTTCATTT AAAAAAAT	6000
	GTGTCTTTTG GCCCTGATT CCCCTCCGTC TCCTCGTGTG GCTGCATTGA AAAGGTTTCC	6060
	TTAGGATGAA AGGAGAGGGG TGTCTCTGT CCCTAGGTGG AGAGAAACAG GGTCTTCTCT	6120
30	TTCTCCGTT TTTTCACCTA CCGTTTCTAT CTCCCTCCTC CCCTCTCCAG CCCTGTCTC	6180
	TGCTACAAAC CACCCCTCC TCCCTCCGGC TGTGGGAGC GCAGGAGCAC GTTGGGCATC	6240
	TGGATGAGCG GNAGACTATT AGCGGGGCAC GGGGGCTCCC CGAGGAGCGC GCGAATTCAC	6300
35	GCTGCCCCAT GAGACCAGGC ACCGGGGGGC GGAGGGGCCT TGGGTGTCCG CAGAGGGACG	6360
	GGCGGGCAGA GCCTTCCTCC GCATTCTAAA CATTCACTTA AAGGTATGAG TTTANTTTCA	6420
	GGGGTGCTGC TGGGAGAGCC TCCAAATGGC TTCTTCCAGC CCCTGCCTGA CAGTTCAGCT	6480
40	CCCCTGGAAG GTCAACTCCT CTAGTCCTTT CTCTTGGTTC TGGGCAGGAC AGAAGTGGG	6540
	GGAGGGAGAG AGAGAGAGAG AGAGAGAGAG ACGGTCAGGA TCCCCGACC CTGGGGAACC	6600
	CGTCAAAAAT AAATGAAATT AAGATTGCCG ACCAGAGAGA GAACCGTGAC AAAGCAAACG	6660
45	GCGTTCAAAG CAAAGAGACG AACTGAAAGC CCGTTCCCGT AGGACTGGTT ATGAGGTCAA	6720
	CACATTCAAA CACAGCTTGC TCTGGATTTT GCTGAGCAGA GGAAGATACA GATGCATTTG	6780
	ATCCAAAGTG TGTTACATCT TTCATTATAT GTGTGTCTAT ATATATAAAC ATATATAAAT	6840
50	ATATAACAT ACATAAATGT ATGTAAATAT ATATAATCTA TATACATATA TAAATATATA	6900
	AACACATATA TAATATATAA ATCTATAAAC ATATATAATA TATAAACATA AATATATAAA	6960
	CATATATAAT ATATAAATAT ATTAACATAT ATAAAATATG TATAAATATA TATAACATA	7020
55	TAAACATATA TAAATATATA AACATATAAA TATATAAACA TATATAAATA TATACAAACA	7080

EP 1 260 228 A2

	TATTGTATAT ATATAAATAT ATATAAAAAC ATATATATAC ATATAAAAAT ATATATAAAC	7140
	ATATATACAT ATAAAGAAAT ATATATAAAC ATATATACAT ATAAATATAC ATATATAAAC	7200
5	ATATATATAC ATAAAATATA TATAAACATA TATACATATA AAAATATATA TATATTAAAC	7260
	TATATATACA TATAAAAATA TATATATTAA CATATATATA CATATAAAAA TATATATATA	7320
	TTTTTGGCCC CTGATTCCCT TCGGTTCCCTG TGGGATGGGT GATTGAGTCA ACACATTCAA	7380
10	ACACAACCTT TCCATCGATG TTGCTTAGGA GATGAGGATA CAGATGCGTT TGATGGAGAG	7440
	GGTTTTACAA GCTCTTTCAT TTAAATATAT ATATATATAT ATATATATTT TTTGGCTCCT	7500
	GATTCTCTTC CGTCTTCCCA TGTGGCTGCA TTTTAAAAGG CTTCCCTAAG ATCGTTACGA	7560
15	TTAAATCAAC CCTCCCCAGG CATCTTTACC GAGGGCTGTG GTCCCCAAG CGATACAGCC	7620
	CAGGAGGGAG AGAGGCTTTG GTGACTTGGA GGAAGGACTG TGTCCCTCCT TAGGGCGTCT	7680
	GTGGCCTCAG TGAGGGAAGG AAGCTGCATC AGACAGGGGT TTCCTCGCTG TCCACCCCTC	7740
20	TGGCAGAAGA TGGATTGGGC TGCCCCGNTA TAAATTAATG AAAAGATTAA AGTTTCGCTA	7800
	AAGGGGACAT CGAGTTTATG TGTATCTCC TGGTGNCTG TGTGCCNTGG GATNCTGCAA	7860
	TATATCCCAN NGCCCTTGAT GNNNTACTGT TTNCTATAAA AANNATAATN TACTTGTNNA	7920
25	ATTTAANTTC CMNNACTA TTTNCTTTCC NNGTNAGTCT NATTANCCGA NCGAGAGCAN	7980
	CGNTTAGTTN CAGCTNGCGG AAAATTGGTT GTGGGGTGTG TGCGGACCCC NGAGNAACGC	8040
	CCNNTAAAAT NAAAGACAAA NTCNNGGGAC AAGNCTNNGG GGTATATCGNN ATTGCNNAGG	8100
30	GGTCGNCATG AAAANTTTAA CGACGGTAAA TAATAATAAA AANNCAAACA TGGGAATGNC	8160
	AATAAAAGAC ATAATCTCC NNTATCGCCG GGGGGGAAAG GATCCTATAG TAAAGGCGAG	8220
	TGCGCTTTGA GGGGTCATAA AAATCAATTA GTTCCAACAC CCACGTCCCG CGTTGAGGGG	8280
35	ACGGGGACGA GCAGGGACAG AAAAAGAAAC CATATTTGAA TCCCATCTCT CTGTGAATTC	8340
	TTGGGTCACA TGCGTCTCAG TACAGCCCGT CCCGTGCTGT GACCGGATAG AGTTTCAATT	8400
	TACTGTGGAA ATTTGCTGTA AATAAATTGA GCATCCGATA GAAGCTGTTG CTGATTAACC	8460
40	TTTTATTTTT AGCGTGGCCC TGCAAAGTCG TATCACCAG CTGTCAGGCT TCTAATCGAA	8520
	AGTTATGAGA CCACGGTGAG GGGCAGGCGG TAATTTAATT ACAACAAATA TCTTTGGGTT	8580
	TATGGCGCAG AGCTAAATTA AATGTCAATTA TTCACTGTCT GTNAATGGNA AATCAAAANN	8640
45	GGAAATCGCA NTTACGNCA TTTGGGNAA ANGAAAGCGG GGNAGTGTCT TTTAATNGAA	8700
	NNGAAATAAC TGTCTTAAGC AGTGTCACAC ACTTCACCTA CCATATTCGN GGCTTNAATT	8760
	GGAANN TGGA TCGTNNGAAT CACTCCNAAG ACTNGATTTA TTANGCGCTT CACGNCAGCN	8820
50	NGGCNTAATT CATCNACTN NGTATCTTC ATCNNNNAT TTTTTTTTC CTCTCNNGCC	8880
	GTGTTNNGAA GGGAGAGTGA ATGAGGCTTT CCACGTTTCA GGAGGATTTT CTTTTTTGAA	8940
	AAATGCCCTT CCAGAGGCTT TTGGGTGGCT GGCTTGCTTT CTGGGCCCTG GAGGANGACA	9000
	GGCGGANGAG TCCAGGTGGG CATGGAGAGG CACAGTGGCA GGTACCTGG ATGGTCAGTG	9060
55	GAGGTGGAGG TCTGAAGGCG CCAGCTTTGG AAATTATTGG TGAATTCGA TGTCAGCACC	9120

EP 1 260 228 A2

	AGGNCAGGGG CCTTTTGGC GGGGGTGTGA GGGANGGATG ANCTTTGCTG GGAAANNACG	9180
	GATCAGGTTC TCCAGGCGCA CTGCAGCCCC GTAGGACCCA CTTTGGAAAT GAAAAGCCAG	9240
5	TTNCCGAAAG CTGGGCTGGA AGCTTCCGTG TTGGGTTCAG GAGCAAGTTC ACGTTGCGCT	9300
	GTGTAGACTC CTGGCTGCTC CCAAACCTCTG AGGGTTTCT GAGGTTCCT TCATAGGGGC	9360
	ACCGGCCCTG GGCCATGCAC AGTGCCTAAG GGTGGCTGTG GGCCGAGGGA CCCAGCACGT	9420
10	GTTTTGCCCA CAACAGCCGG AGTGACTGGT TCACTCACC GCTTGGCGGA GGACGCCTGT	9480
	TCTCTGGACG AATCATTTCT CTTGGGTGGT GACTGCCTTG TGGGTCAAGG TGCAGGTTTT	9540
	CTGCCACAGA AAACCTGTTA GGAGGAATTA AGCGACTAAG ACTGTCAGG AGGTGGTGGT	9600
15	GGGGGANGAG GNAGGGGGTG GTGTCCAGAT TACCAGGCAT AGGCTAAACT GCCTGCACTC	9660
	TCCAGCTGGT CTGTCTGTGG AGGAGGGGAT TGTCAATACT GGGAGAGCAG AGGAGGCTCG	9720
	TAGGAGGTGA GAGGGGGTGG AATTGTGCATG CAAATCTTCA CATGAGGCC TGTGAATTT	9780
20	CTCCAGCCTC CTGAGGGTCC CTTGCGCTAT TGCCTCAAC TTCTTGATAG TTTACCCCAA	9840
	GACTCAGAAG TCCTTAGAGG GGCAGAATGC CCCCACCACA AAGCCTGCTA TCCTTGGGCG	9900
	TCCTCAGGAC CTTGGTTCAT GAATGGGACC CTTTCATGTA TGGGGACCCT TGGTAATATG	9960
25	AATGGGACGC CTTCAGCTCC CCAGGGCTTC CGAGGAGGCC GAGAAGGGCA AAGACACTTC	10020
	CGAGGAGGCC GAGAAGGGCA AAGACATTTT CTGGGCTTGG TGTGTCAAGA GCTAGATTGG	10080
	AGAAGGGGCT GGATTGGAA CTCTTTAGCC ATCAGCTCAC CCTCTCCGTT TGTGGCTAAA	10140
30	GTCTGAAGGT GGAAACTTCG GTTCTCTAC AGGTCTACA GGAGTTGGGG GCGGGGCGC	10200
	CCACACAGAA CGCTGGAAAG TTCGACAGTC CACTTCCACT GGCTCGGAAC TCCTTTTTTC	10260
	ACCTTAAGTT CATCAGCGGT AACGCATAGG TCTCACTTAG GCAGGGCACG GATGATTTAA	10320
35	CAATTTCTAC TTCTAGGTCA GGTGCGGTGG CTCACACCTC TAATCCCAGC ACTTTGGGAG	10380
	GCCCAGGAGG GTGGATCGCT TGAGGTGAGG AGTTTGAGAC CAGCCTGGCC AACATGGTGA	10440
	AACCCCGTCT CTAATAAAAT ACGAAAATTA GCCAGGCATG GTGGTGAGCA CCTGTAATTC	10500
40	CAGCTACTCG GGAGGCTGAG GCAGGAGAAT CGCTTGAACC TGGGAGGTGG ACGTTGCAGT	10560
	GAGGTGAGAT CACACCACTG CACTCCAGCC TGGATGAGAG AGCAAGACTC TGTCTCAAAA	10620
	ACAAAATAAA AAAAAACAA AAAAAAATC AAAAAAGAAA ACCCAATTTT CAGTTCTAGG	10680
45	CCAGGTGCAG TGGCTCACGC CTGTCAATCC AGCACTTTGG GAGGCCCAGG AGGTGGATC	10740
	GCTTGAGGTC AGGAGTTTCA GACCAGCCTG GCCAACATGG TGAAACCCCA TCTTTACTAA	10800
	AAATACAAAC GTTAGCTGGG TGTGGTGGTG TGCGCCTGTA ATCCCAGCTA CTCGGGAAGC	10860
50	TGAGGCTGGA GAATTGCTTG AATCTGGGAG GTGGAGGTTG CAGGGAGGCG AGATAGTGCC	10920
	ACTGCAGTCC AGCCTGGACC AGAGAGCAAG ACTCCGTCTC AAAAAACAAA GAAAGCAAAA	10980
	ACAAAAACA AGAGACCAGC CTGGCCAACA TGGTGAAACC GCGTCTTTAC TAAAAACAA	11040
55	AATTAGCCGG GCATGGTGGT GGGCACCTGT AGTCCAGCT ACTCGGGAGG CTGAGGCAGG	11100

EP 1 260 228 A2

	AGAATGGCTT GAACCTGGGA GGTGGAGCTT GCAGTGAGCC GAGATAGTGC CACTGCACTC	11160
	CAGCCTGGGC GACAGAGCGA GACTTGATTT CAGAACCACC ACCACCACAA CAAAACAAAA	11220
5	CAAAAAATCC AAAAAAACCC CAATTTCCAG TACTAGGTAG TCAGTGATGC AGGGCTGGAG	11280
	ACAGAGGGGC GGTAAGTGTC TGGGCGCCCA CCATCAGTCA CCTCCCAGCT CCCANGAGGT	11340
	GCAAAGTGCT TGGTTCAGCC TCATGGGAAG GATGCTCCCT GGGGAGGCTG GGCTGGGTTC	11400
10	ACAGGGCTCT TCACATCTCT CTCTGCTTCT NCCCCAAGGT TTGGTTNCCA GAACCGGAGA	11460
	GCCAAGTGCC GNCAACAAG AGAATCAGAT GCATAAAGGT GGGTGTGGG ACTGGGGGGA	11520
	CCTGAAGCTG GGGGATCCTG CTCCAGGAGG GATGGGGTCG ACAAGGTGCT GGCTACACCC	11580
15	AGGACCACCA CACTGACACC TGCTCCCTTT GGACACAGGC GTCATCTTGG GCACAGCCAA	11640
	CCACCTAGAC GCCTGCCNGA GTGGCACCTT ACGTCAACAT GGGAGCCTTA CGGATGCCTT	11700
	TCCAACAGGT AGCTCACTTT TTCTTCCTCT GNAAGATCCC TAGGGACCTG CTGCTCCCTT	11760
20	CCCCTTTCCC CTATTTGCTG CCGCATCCTG ACACTCCTAG TCCCTCCCTG CCCCTGCAGA	11820
	CTTCTCAGCT GGCCCTTAGA AAAAAAGCCT CTTTTCGAG GAGGCATTTA CAGGCACCTT	11880
	GGCACCTATG AAATCAGGCT GGGCCAGGCG GGGTGGCTCA CACCTGTCTAT CCCAGCACTT	11940
25	TGGGAGGCCA AGGTTAGGAG TTTGAGACCA GCCTGGACAA CATAGCAAAA GCCTGTCTCT	12000
	ACTAAAAATA CAAAAAATAA TTAACAGGGA GTGGTGGTGG GCACCTGTAA TCCCAGCTAC	12060
	TTGGGAGGCT GAGGCAGGAG AATCACTTGA ACCCGGGAGG CCGAGGTTGC GGTGAGCCGA	12120
30	GATCGTGCCA TTGCACTCCA GGCTGGGCGA CAGAGTGAGA CTCTGTCTCA AAAAATAAAT	12180
	AAATAAATAA ATGTAAAAAA ATAAAAATAG GTCGGGCACG GTGGCTCACG TCTGTAATCC	12240
	CAGCACTTTG GAAGGCCGAG GTGGGTGGAT GACAGGGTCA AGAGATTGAG ACCATCCTGG	12300
35	CCAACATGGC AAAATGCCGT CTCTACTAAA AAATACAAAA ATTAGGCGGG CGTGGTGGCG	12360
	GGTGCCTGTA ATCCCAGCTA CTCGGGAGGC TGAGGCAGGA GAATCGGTTG AACC CGGAT	12420
	GCGGAGGTTG CAGTGAGCGG AGATCACATC ACTGCACTCC AGGCTGGGCA ACAAGAGCGA	12480
40	AACTGCGTCT TACAATAAAT AAATAGATAA ATAAATAAAC AAATAAACTT TACTTTAGAA	12540
	ACAAATCCCT GTCCGTGTTT GTCTTTTCAC CTGTCTTGCA GGGAAAACAA AACATAAAAT	12600
	GTCAAGGCAA ATAGTAGTGA TTTCAATCCG GGAAAAAGAA AGTGGATGTT TGCCTTCACC	12660
45	CTTTCTCGTC CTTCTCTTGG TGCTCCTCAN GGCCCANGGG NAGAGGGTGG AAAGTNCAGA	12720
	GGAAGAAAGA CGGGGCTGGG GGGGGGGGTC CGTGGGGACC CAGGCAGGCA TGTTCNAT	12780
	TTCCNTGTCT TCACNTTCAA AGNAGGGGCC CCTCGNCTCT GGAATGAGGC CTACGGTTTC	12840
50	CTTTCCNCA AGAGTTNCCC CTTTGTGAGC TTACGGCTTC GGAGTGAACC TCGGTGCAAC	12900
	CTGTTATTA AACAACAGAG GGCTAATGCC AGCAAAAACA CGCCCCCGC TCCTGGTTTC	12960
	AGAGGGAAGA AAAAAATTC TAAGCACGGC CATGCTTTTC TAATAAAAAAT TCATTAAATA	13020
	ATCGTTATAA GGGATGAAGC CGGGAGGGGA GAGGAGAGGA ACACAATCAA GAGACTTTCT	13080
55	TTGAACTTT TCTCCCTGCT TCAAATACAA AGCAATCTTC TGTGGGCCTG GGCTGGGGG	13140

EP 1 260 228 A2

	GTTTCCCCCT TTCTCTGCAG CCCATTGGGA GGAAGAAAAT GCTTCCCTGA ANGTTGCTGC	13200
	AAAATTGTTT CTGTTTTFCT TTTCTTTTTC TTTTMTTTTT TTTTGTGAGA CGGAGTCTCG	13260
5	CTCTGTCACC AGGCTGGAGT GCAATGGTAT GATCTCAGCT CACTGCAACC TCCACGTTC	13320
	TGTTTCAAGT CATTCTCCTG CCTCAGCCTC CTGAGTAGCT GGGACTACAG GCGCCCGCCA	13380
	CCACGCCCCG CTAGTGTTTG TATTTTTAGA AAAGACAGGG TTTCCCATG TTGCCAGGC	13440
10	TGGTCTTGAA CTCCTGTCCT CAAGTGATCT GCCTGCCTCG GCCTCCCAA GTGCTGTGTT	13500
	TCTGTMTTTC TTTCCCGCT TTCTTAGGAG GCCATCGGA AGAATAAAAT GCTTTCCTTG	13560
	AAGTTGATGC AAAATTGTTT CTGTTTTTCT TTTCTCTTTT CTTTCTTTTT GAGATGGAGT	13620
15	CTCGCTCTTT CACCCAGGCT GGAGGGCAGT GCGCGACCT CGGCTCACTG CAACCTCCGC	13680
	CTCCCGGGTT CAAGCGATTC TCCTGCC TCA GCCTCCGAG TAGCTGGGAT TACAGGCACC	13740
	TGCCACTATG CCTGGCTAAT TTTATTATTT TTAGTAGAGA CGGGGTTTCA CCATGTTGGC	13800
20	CAGGCTGGTC TCAAACCTCT GACCTCAGGT GATCCGCCCG CCTCGCCTCC CAAAGTGATG	13860
	GGATGANCAG GNCAATNGAG NCACCGTGCC CGGCCCTCTA ACTCTTTACC AGACATAAAG	13920
	TCTCCNNTTC CCCTTTCTAA ATGTATATAT TGTGTTTTTA AAAGTTAACA GCAGGGATCC	13980
25	CACCTCATTN CCCCCTNCT CTCCCCAAGA CCTGTCTTGC ACGTTGCACA CAGCAGGTGT	14040
	GCCCTGGACA TATCCCAAAC CCACGCTGAA AGAAAGAGGG TCTCACTACA CGTATGATAT	14100
	CTGTGNATCC TTTAAACATC TCCGTGGCTT CCAGGCAACA CAGCCATAAA TAGGAATCTC	14160
30	ATGTCTGACA TGATACCGG ACCATGTATG GGNAAATTCT GGGTGTGAAG TTCCAGCTAC	14220
	CCCCGCAGAG GCANCCATTG CATACCCTCC AGAAACTCCC CTGCCGTTNC AAGCCAAAGA	14280
	CACAACACAA ACAGCNTCCG AGAGAGGGTG TCATTGAAAA TCAATACCAT CATAAGAGCA	14340
35	CACAGCACCG TCTTTCTCTT CTGCCCCGTTG ATACACAATT ATGAGCAATT TGCTAACACT	14400
	GACAACTCGT GGCAAGAACA GGTGCTGTTG ATACGGTTGC CTCGTGAGGA CCCATCTGTC	14460
	TTCTGGGGTC TTGCCTGGAA CGGAGATCGG AGTTCAGGGT GGCTAATAGA ATCATTACTC	14520
40	ACCTAGGGAC ACAGAATNAT GAGGGTTACC CCCAGTTAAG TGCATACAGT CAAACGGACG	14580
	GCTGCTCTGG AAGGTACAGT GACGTGAACA GCTTTTATGA AATGCCCTAGA TCTGGACCTT	14640
	CCATACCTGA GCCACCGTTC CAAAGCACTG GCGGTTTTTC AGATACTTTC ATGAGAAATG	14700
45	TTGTCAACAC CGCAAGTTTG CAGTACACAG TCTGAAAGAT ATTCTTGTAT ATGTAGATGT	14760
	CTGTAGATGC CCTGAAGGTG TGTAGACTTT AGACACCCAG AAGGTGTGTA GATGTCTGTA	14820
	GACACCTTCT ATGTGTGTAG ATGTCTGTAG ACGCCCTGCA GGTGTGTAGA TATATCTAGA	14880
50	TGGTCTGCCT GTGTATGATA CAGGCTAAAA AGACATTTGT GGTGGACACT AGTTGATTAT	14940
	TTAGGACTAT GAGATGGGAA AGGAAGNAGC AACCAGCAGT GAAAGGCATG TGGTGGGTGG	15000
	GGGGTTGGCA TTGCAGTGGG GTCCTCNTGA NGCAGGTGAC ACCCACTATA GGGCTGCCCT	15060
55	TGGNATGGAC GCTTTGTNGA AGCTGTTTGA TTTCACCACA CCAAGCCTGG AGGCACGGAC	15120

EP 1 260 228 A2

ATTCCAGGAT GGTGAGGAGT CTGCAAAGGA GGAGATTGGA GGAGGTGCAA TATCCCTAGA 15180
 GTACGAGAGA TGAGATAGGA GAGCTGTATA AATAGCACTA CCAGCCGGAT GCGGTGGCTC 15240
 5 ACGCCTGTCA TCCCAGCACT TTAGGAGGCT GAGGCAGGCG GATCACCTGA GGTCAGGAGT 15300
 TCCAGAACAG CCTGGCCAAC ACAATGAAAC CCCATCTTTA CTAAAAATAC AAGATTAGCT 15360
 GGGCACGGTG TCTCACGCCT GTCATCCCTG CACTTTGGGA GGTGAGGTG CGCAGATCAT 15420
 10 GAGGTCAGTT TGGCCAACGC GGCGAAACCC CGTCTCTACT AAAAATACAA AAAAGTAGCC 15480
 GGGCGTGGTG GTGGGCACCT GTAGTCCCAG CTACTAGGGA GGCTGAGGCA GGAGAATCGC 15540
 TTGAACCCGG ATGCGGACAT TGCAGTGAGC CGAGATC 15577

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 753 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "ET92 gene segment"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CGTGGAAGCC TGGAGTTTTT GGGAACAGCG TGTCCCCGCC GAGCCTGGGA GCCCGTGGGT 60
 30 TCTGCAAAGC CTGCGGGTGT TTGAGGACTT TGAAGACCAG TTTGTGAGTT GGGCTCAATT 120
 CCTGGGGTTC AGACTTAGAG AAATGAAGGA GGGAGAGCTG GGGTCGTCTC CAGGAAACGA 180
 TTCACTTGGG GGGAAGGAAT GGAGTGTCTT TGCAGGCACA TGTCTGTTAG GAGGTGAAAC 240
 35 AGAATGTGAA ATCCACGTTG GAGTAAGCGT CCAGCGCTGA ATGTAGCTCG GGGTGGGGTG 300
 GGAGGGCCCT GGTGTGGATC GTGGAAGGAA GAAAGACAGA ACAGGGTGCT AGTATTTACC 360
 CCGTTCCCTG TAGACACCCT GGATTGTGCA GCTTTGCAAG CTTCTTGGTT GCAGCGGCCT 420
 40 TGCCTGTGCC CTTTGTAGAC TGTTTCCAGA CTAACTTCC AAATGTCAGC CCCTTACCCT 480
 TGACAGCAAG GGACATCTCA TTAGGCATC GCGTGCTTCT CATCTGTGCT CAGCAGGCCC 540
 GAGATAGGAA CAGAGGGGCG TTGGAGATGC CACTTCCACC AGCCCTGGGT TGAAGGGGAG 600
 CGAGGGAGAC ACCTTTTACT TAAACCCCTG AGCTTGGTCA GAGAGGCTGA ATGTCTAAAA 660
 45 TGAGGAAGAA AAGGTTTTTC ACCTGGAAC GCTTGAGGGC TGAGTCTTCT GCCCTTCTGA 720
 CTCCCCCAGC AAATACAGAC AGGTCACCAA CTA 753

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1890 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

EP 1 260 228 A2

(A) DESCRIPTION: /desc = "SHOXa"

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:91..968

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

10	GTGATCCACC CGCCGCACGG GCCGTCTCTT CCGCGCGGGG AGACGCGCGC ATCCACCAGC	60
	CCCGGCTGCT CGCCAGCCCC GGCCCCAGCC ATG GAA GAG CTC ACG GCT TTT GTA	114
	Met Glu Glu Leu Thr Ala Phe Val	
	1 5	
15	TCC AAG TCT TTT GAC CAG AAA AGC AAG GAC GGT AAC GGC GGA GGC GGA	162
	Ser Lys Ser Phe Asp Gln Lys Ser Lys Asp Gly Asn Gly Gly Gly Gly	
	10 15 20	
	GGC GGC GGA GGT AAG AAG GAT TCC ATT ACG TAC CGG GAA GTT TTG GAG	210
20	Gly Gly Gly Gly Lys Lys Asp Ser Ile Thr Tyr Arg Glu Val Leu Glu	
	25 30 35 40	
	AGC GGA CTG GCG CGC TCC CGG GAG CTG GGG ACG TCG GAT TCC AGC CTC	258
	Ser Gly Leu Ala Arg Ser Arg Glu Leu Gly Thr Ser Asp Ser Ser Leu	
	45 50 55	
25	CAG GAC ATC ACG GAG GGC GGC GGC CAC TGC CCG GTG CAT TTG TTC AAG	306
	Gln Asp Ile Thr Glu Gly Gly Gly His Cys Pro Val His Leu Phe Lys	
	60 65 70	
	GAC CAC GTA GAC AAT GAC AAG GAG AAA CTG AAA GAA TTC GGC ACC GCG	354
30	Asp His Val Asp Asn Asp Lys Glu Lys Leu Lys Glu Phe Gly Thr Ala	
	75 80 85	
	AGA GTG GCA GAA GGG ATT TAT GAA TGC AAA GAG AAG CGC GAG GAC GTG	402
	Arg Val Ala Glu Gly Ile Tyr Glu Cys Lys Glu Lys Arg Glu Asp Val	
	90 95 100	
35	AAG TCG GAG GAC GAG GAC GGG CAG ACC AAG CTG AAA CAG AGG CGC AGC	450
	Lys Ser Glu Asp Glu Asp Gly Gln Thr Lys Leu Lys Gln Arg Arg Ser	
	105 110 115 120	
	CGC ACC AAC TTC ACG CTG GAG CAG CTG AAC GAG CTC GAG CGA CTC TTC	498
40	Arg Thr Asn Phe Thr Leu Glu Gln Leu Asn Glu Leu Glu Arg Leu Phe	
	125 130 135	
	GAC GAG ACC CAT TAC CCC GAC GCC TTC ATG CGC GAG GAG CTC AGC CAG	546
	Asp Glu Thr His Tyr Pro Asp Ala Phe Met Arg Glu Glu Leu Ser Gln	
	140 145 150	
45	CGC CTG GGG CTC TCC GAG GCG CGC GTG CAG GTT TGG TTC CAG AAC CGG	594
	Arg Leu Gly Leu Ser Glu Ala Arg Val Gln Val Trp Phe Gln Asn Arg	
	155 160 165	
	AGA GCC AAG TGC CGC AAA CAA GAG AAT CAG ATG CAT AAA GGC GTC ATC	642
50	Arg Ala Lys Cys Arg Lys Gln Glu Asn Gln Met His Lys Gly Val Ile	
	170 175 180	
	TTG GGC ACA GCC AAC CAC CTA GAC GCC TGC CGA GTG GCA CCC TAC GTC	690
	Leu Gly Thr Ala Asn His Leu Asp Ala Cys Arg Val Ala Pro Tyr Val	
	185 190 195 200	
55	AAC ATG GGA GCC TTA CGG ATG CCT TTC CAA CAG GTC CAG GCT CAG CTG	738
	Asn Met Gly Ala Leu Arg Met Pro Phe Gln Gln Val Gln Ala Gln Leu	

EP 1 260 228 A2

	205	210	215	
5	CAG CTG GAA GGC GTG GCC CAC GCG CAC CCG CAC CTG CAC CCG CAC CTG Gln Leu Glu Gly Val Ala His Ala His Pro His Leu His Pro His Leu 220 225 230			786
	GCG GCG CAC GCG CCC TAC CTG ATG TTC CCC CCG CCG CCC TTC GGG CTG Ala Ala His Ala Pro Tyr Leu Met Phe Pro Pro Pro Pro Phe Gly Leu 235 240 245			834
10	CCC ATC GCG TCG CTG GCC GAG TCC GCC TCG GCC GCC GCC GTG GTC GCC Pro Ile Ala Ser Leu Ala Glu Ser Ala Ser Ala Ala Ala Val Val Ala 250 255 260			882
15	GCC GCC GCC AAA AGC AAC AGC AAG AAT TCC AGC ATC GCC GAC CTG CGG Ala Ala Ala Lys Ser Asn Ser Lys Asn Ser Ser Ile Ala Asp Leu Arg 265 270 275 280			930
	CTC AAG GCG CGG AAG CAC GCG GAG GCC CTG GGG CTC TG ACCCGCCGCG Leu Lys Ala Arg Lys His Ala Glu Ala Leu Gly Leu 285 290			978
20	CAGCCCCCGG CGCGCCCGGA CTCCCGGGCT CCGCGCACCC CGCCTGCACC GCGCGTCCTG			1038
	CACTCAACCC CGCCTGGAGC TCCTTCCGCG GCCACCGTGC TCCGGGCACC CCGGGAGCTC			1098
	CTGCAAGAGG CCTGAGGAGG GAGGCTCCCG GGACCGTCCA CGCACGACCC AGCCAGACCC			1158
25	TCGCGGAGAT GGTGCAGAAG GCGGAGCGGG TGAGCGGCCG TCGCTCCAGC CCGGGCCTCT			1218
	CCAAGGCTGC CCGTGCGTCC TGGGACCCTG GAGAAGGGTA AACCCCGGCC TGGCTGCGTC			1278
	TTCTCTGCT ATACCTATG CATGCGGTTA ACTACACAG TTTGGAAGAT CCTTAGAGTC			1338
30	TATTGAACT GCAAAGATCC CGGAGCTGGT CTCCGATGAA AATGCCATTT CTTCGTGACC			1398
	AACGATTTTC TTTACTACCA TGCTCCTTCC TTCATCCCGA GAGGCTGCGG AACGGGTGTG			1458
	GATTTGAATG TGGACTTCGG AATCCCAGGA GGCAGGGGCC GGGCTCTCCT CCACCGCTCC			1518
35	CCCGGAGCCT CCCAGGCAGC AATAAGGAAA TAGTTCTCTG GCTGAGGCTG AGGACGTGAA			1578
	CCGCGGGGCTT TGGAAAGGGA GGGGAGGGAG ACCCGAACCT CCCACGTTGG GACTCCCACG			1638
	TTCCGGGGAC CTGAATGAGG ACCGACTTTA TAACTTTTCC AGTGTTTGAT TCCCAAATTG			1698
40	GGTCTGGTTT TGTTTTGGAT TGGTATTTTT TTTTTTTTTT TTTTTTGCTG TGTTACAGGA			1758
	TTCAGACGCA AAAGACTTGC ATAAGAGACG GACCGTGGT TGCAAGGTGT CATACTGATA			1818
	TGCAGCATTA ACTTTACTGA CATGGAGTGA AGTGCAATAT TATAAATATT ATAGATTAAA			1878
45	AAAAAATAG CA			1890

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 292 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Met Glu Glu Leu Thr Ala Phe Val Ser Lys Ser Phe Asp Gln Lys Ser
1 5 10 15

EP 1 260 228 A2

5 Lys Asp Gly Asn Gly Gly Gly Gly Gly Gly Gly Lys Lys Asp Ser
 20 25 30
 Ile Thr Tyr Arg Glu Val Leu Glu Ser Gly Leu Ala Arg Ser Arg Glu
 35 40 45
 Leu Gly Thr Ser Asp Ser Ser Leu Gln Asp Ile Thr Glu Gly Gly Gly
 50 55 60
 10 His Cys Pro Val His Leu Phe Lys Asp His Val Asp Asn Asp Lys Glu
 65 70 75 80
 Lys Leu Lys Glu Phe Gly Thr Ala Arg Val Ala Glu Gly Ile Tyr Glu
 85 90 95
 15 Cys Lys Glu Lys Arg Glu Asp Val Lys Ser Glu Asp Glu Asp Gly Gln
 100 105 110
 Thr Lys Leu Lys Gln Arg Arg Ser Arg Thr Asn Phe Thr Leu Glu Gln
 115 120 125
 20 Leu Asn Glu Leu Glu Arg Leu Phe Asp Glu Thr His Tyr Pro Asp Ala
 130 135 140
 Phe Met Arg Glu Glu Leu Ser Gln Arg Leu Gly Leu Ser Glu Ala Arg
 145 150 155 160
 25 Val Gln Val Trp Phe Gln Asn Arg Arg Ala Lys Cys Arg Lys Gln Glu
 165 170 175
 Asn Gln Met His Lys Gly Val Ile Leu Gly Thr Ala Asn His Leu Asp
 180 185 190
 30 Ala Cys Arg Val Ala Pro Tyr Val Asn Met Gly Ala Leu Arg Met Pro
 195 200 205
 Phe Gln Gln Val Gln Ala Gln Leu Gln Leu Glu Gly Val Ala His Ala
 210 215 220
 35 His Pro His Leu His Pro His Leu Ala Ala His Ala Pro Tyr Leu Met
 225 230 235 240
 Phe Pro Pro Pro Pro Phe Gly Leu Pro Ile Ala Ser Leu Ala Glu Ser
 245 250 255
 40 Ala Ser Ala Ala Ala Val Val Ala Ala Ala Ala Lys Ser Asn Ser Lys
 260 265 270
 Asn Ser Ser Ile Ala Asp Leu Arg Leu Lys Ala Arg Lys His Ala Glu
 275 280 285
 45 Ala Leu Gly Leu
 290

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1354 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "SHOXb"

EP 1 260 228 A2

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 91..768

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

	GTGATCCACC CGCCGCACGG GCCGTCCTCT CCGCGCGGGG AGACGCGCGC ATCCACCAGC	60
10	CCCGGCTGCT CGCCAGCCCC GGCCCCAGCC ATG GAA GAG CTC ACG GCT TTT GTA Met Glu Glu Leu Thr Ala Phe Val 295 300	114
15	TCC AAG TCT TTT GAC CAG AAA AGC AAG GAC GGT AAC GGC GGA GGC GGA Ser Lys Ser Phe Asp Gln Lys Ser Lys Asp Gly Asn Gly Gly Gly Gly 305 310 315	162
20	GGC GGC GGA GGT AAG AAG GAT TCC ATT ACG TAC CGG GAA GTT TTG GAG Gly Gly Gly Gly Lys Lys Asp Ser Ile Thr Tyr Arg Glu Val Leu Glu 320 325 330	210
25	AGC GGA CTG GCG CGC TCC CGG GAG CTG GGG ACG TCG GAT TCC AGC CTC Ser Gly Leu Ala Arg Ser Arg Glu Leu Gly Thr Ser Asp Ser Ser Leu 335 340 345	258
30	CAG GAC ATC ACG GAG GGC GGC GGC CAC TGC CCG GTG CAT TTG TTC AAG Gln Asp Ile Thr Glu Gly Gly Gly His Cys Pro Val His Leu Phe Lys 350 355 360	306
35	GAC CAC GTA GAC AAT GAC AAG GAG AAA CTG AAA GAA TTC GGC ACC GCG Asp His Val Asp Asn Asp Lys Glu Lys Leu Lys Glu Phe Gly Thr Ala 365 370 375 380	354
40	AGA GTG GCA GAA GGG ATT TAT GAA TGC AAA GAG AAG CGC GAG GAC GTG Arg Val Ala Glu Gly Ile Tyr Glu Cys Lys Glu Lys Arg Glu Asp Val 385 390 395	402
45	AAG TCG GAG GAC GAG GAC GGG CAG ACC AAG CTG AAA CAG AGG CGC AGC Lys Ser Glu Asp Glu Asp Gly Gln Thr Lys Leu Lys Lys Arg Arg Ser 400 405 410	450
50	CGC ACC AAC TTC ACG CTG GAG CAG CTG AAC GAG CTC GAG CGA CTC TTC Arg Thr Asn Phe Thr Leu Glu Gln Leu Asn Glu Leu Glu Arg Leu Phe 415 420 425	498
55	GAC GAG ACC CAT TAC CCC GAC GCC TTC ATG CGC GAG GAG CTC AGC CAG Asp Glu Thr His Tyr Pro Asp Ala Phe Met Arg Glu Glu Leu Ser Gln 430 435 440	546
60	CGC CTG GGG CTC TCC GAG GCG CGC GTG CAG GTT TGG TTC CAG AAC CGG Arg Leu Gly Leu Ser Glu Ala Arg Val Gln Val Trp Phe Gln Asn Arg 445 450 455 460	594
65	AGA GCC AAG TGC CGC AAA CAA GAG AAT CAG ATG CAT AAA GGC GTC ATC Arg Ala Lys Cys Arg Lys Gln Glu Asn Gln Met His Lys Gly Val Ile 465 470 475	642
70	TTG GGC ACA GCC AAC CAC CTA GAC GCC TGC CGA GTG GCA CCC TAC GTC Leu Gly Thr Ala Asn His Leu Asp Ala Cys Arg Val Ala Pro Tyr Val 480 485 490	690
75	AAC ATG GGA GCC TTA CGG ATG CCT TTC CAA CAG ATG GAG TTT TGC TCT Asn Met Gly Ala Leu Arg Met Pro Phe Gln Gln Met Glu Phe Cys Ser 495 500 505	738
80	TGT CGC CCA GGC TGG AGT ATA ATG GCA TGA TCTCGACTCA CTGCAACCTC Cys Arg Pro Gly Trp Ser Ile Met Ala *	788

EP 1 260 228 A2

510 515

CGCCTCCCGA GTTCAAGCGA TTCTCCTGCC TCAGCCTCCC GAGTAGCTGG GATTACAGGT 848

5 GCCCACCACC ATGTCAAGAT AATGTTTGT TTTTCAGTAG AGATGGGGTT TGACCATGTT 908

GGCCAGGCTG GTCTCGAACT CCTGACCTCA GGTGATCCAC CCGCCTTAGC CTCCCAAAGT 968

GCTGGGATGA CAGGCGTGAG CCCCTGCGCC CGGCCTTTGT AACTTTATTT TTAATTTTTT 1028

10 TTTTTTTTAA AGAAAGACAG AGTCTTGCTC TGTCACCCAG GCTGGAGCAC ACTGGTGCGA 1088

TCATAGCTCA CTGCAGCCTC AACTCCTGG GCTCAAGCAA TCCTCCACC TCAGCCTCCT 1148

GAGTAGCTGG GACTACAGGC ACCCACCACC ACACCCAGCT AATTTTTTTG ATTTTACTA 1208

15 GAGACGGGAT CTGCTTTGTC TGCTGAGGCT GGTCTTGAGC TCCTGAGCTC CAAAGATCCT 1268

CTCACCTCCA CCTCCCAAAG TGTTAGAATT ACAAGCATGA ACCACTGCCC GTGGTCTCCA 1328

AAAAAAGGAC TGTTACGTGG AAAAAA 1354

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 226 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Glu Glu Leu Thr Ala Phe Val Ser Lys Ser Phe Asp Gln Lys Ser
 1 5 10 15

Lys Asp Gly Asn Gly Gly Gly Gly Gly Gly Gly Lys Lys Asp Ser
 20 25 30

Ile Thr Tyr Arg Glu Val Leu Glu Ser Gly Leu Ala Arg Ser Arg Glu
 35 40 45

Leu Gly Thr Ser Asp Ser Ser Leu Gln Asp Ile Thr Glu Gly Gly Gly
 50 55 60

His Cys Pro Val His Leu Phe Lys Asp His Val Asp Asn Asp Lys Glu
 65 70 75 80

Lys Leu Lys Glu Phe Gly Thr Ala Arg Val Ala Glu Gly Ile Tyr Glu
 85 90 95

Cys Lys Glu Lys Arg Glu Asp Val Lys Ser Glu Asp Glu Asp Gly Gln
 100 105 110

Thr Lys Leu Lys Gln Arg Arg Ser Arg Thr Asn Phe Thr Leu Glu Gln
 115 120 125

Leu Asn Glu Leu Glu Arg Leu Phe Asp Glu Thr His Tyr Pro Asp Ala
 130 135 140

Phe Met Arg Glu Glu Leu Ser Gln Arg Leu Gly Leu Ser Glu Ala Arg
 145 150 155 160

Val Gln Val Trp Phe Gln Asn Arg Arg Ala Lys Cys Arg Lys Gln Glu
 165 170 175

Asn Gln Met His Lys Gly Val Ile Leu Gly Thr Ala Asn His Leu Asp

EP 1 260 228 A2

180 185 190

Ala Cys Arg Val Ala Pro Tyr Val Asn Met Gly Ala Leu Arg Met Pro
195 200 205

5 Phe Gln Gln Met Glu Phe Cys Ser Cys Arg Pro Gly Trp Ser Ile Met
210 215 220

Ala *
225

10 (2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32367 base pairs
(B) TYPE: nucleic acid
15 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = 'COSMID: LLNOYCO3'M'34F5'

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TTTCTCTGTC TCCATCCCTC TGTCTCTCCC TTTCTCTCTG TCTTTCTCTG TCTCTCTCTT 50

25 TCTCTCTCTC TCTCCATCTC TCTCTCTCCC TGTCTCTCTC TCTCCATCTC CCCGTCTCTC 120

CGTTTCTCTC TCTGCCTCTC CCTGTCTGTC TCTCTCTTTC TGTGTCTTAC ACACACCCCA 180

ACCCACCGTC ACTCATGTCC CCCCCTGCT GTGCCATCTC ACACAAGTTC ACAGCTCAGC 240

30 TGTCACTCTG GGTCCCCAGG CCCCCTCGGG GAGGAAGATG CGCCGTGGGG TTACGGGAGG 300

AAGGGGACTC CGGGCTCCT GGTGCCCCAC TTTATTGCA GAAGGTCCTT GGCAGGAACC 360

GTGACGCGTT TGGTTTCCAG GACTTGAAA ACGAATTTCA GGTGCGATG GCGAGCACCG 420

35 GCTTCCCCTG AAGCACATTC AATAGCGAGA GCGGGAGGG AGCGAGCAGG AGCATCCAC 480

CATGAAAACC AAAACACAA GTATTTTTTT CACCCGGTAA ATACCCAGA CGCCAGGGTG 540

ACAGCGCGGC GCTAAGGGAG GAGGCCTCGC GCCGGGTCC GCCGGATCT GCGCGGGCG 600

40 GAAAGAATAT AGATCTTTAC GAACCGATC TCCCGGGAC CTGGGCTTCT TTCTGCGGGC 660

GCTGGAGACC CGGGAGGCGG CCCCCTGGAT CCTCGGCTC CGCCGCGCC GCCTCCAAG 720

CGCCCGCTC CCGTTTGGG GACACCCGCG CCTTCTTCT CACTTTCGGG GATTCTCCAG 780

45 CCGCGTTCCA TCTACCAAC TCTCCATCCA AGGGCGCGCC GCCACCAACT TGGAGCTCAT 840

CTTCTCCCAA GATCGTGCGT CCCCCTGGCG CCCCCTGCC CCCCCTGCC ATCTCAACCC 900

CGGCGCGACC CGGGCGCTC CTGGAAAGAT CCAGCGCCG GGCTCTGCGC TCCTCCCGGG 960

50 AGCGAGGGCG GCCGACGAC TGGGACCTC CTCTCTCCAG CCGTGAAGTC CTTGTCTCTC 1020

TGTCTCTCTC TGCAGGAAAA CTGGAGTTTG CTTTCTCTCC GGCCACGGAG AGAACGCGGG 1080

TAACCTGTGT GGGGGGCTCG GCGCGCTGCG CCCCCTCCT GCGCGCGCGC TCTCCCTTCC 1140

55 AAAAATGGGA TCTTTCCCCC TTCGCACCAA GGTGTACGGA CGCCAAACAG TGATGAAATG 1200

EP 1 260 228 A2

	AGAAGAAAGC CAATTGCCGG CCTGGGGGGT GGGGGAGACA CAGCGTCTCT GCGTGCGTCC	1260
	GCCGCGGAGC CCGGAGACCA GTAATTGCAC CAGACAGGCA GCGCATGGGG GGCTGGGCGA	1320
5	GGTCGCCGCG TATAAATAGT GAGATTTCCA ATGGAAGGC GTAAATAACA GCGCTGGTGA	1380
	TCCACCCGCG CGCACGGGCC GTCCTCTCCG CGCGGGGAGA CGCGCGCATC CACCAGCCCC	1440
	GGCTGCTCGC CAGCCCCGGC CCCAGCCATG GAAGAGCTCA CGGCTTTTGT ATCCAAGTCT	1500
10	TTTGACCAGA AAAGCAAGGA CGGTAACGGC GGAGGCGGAG GCGGCGGAGG TAAGAAGGAT	1560
	TCCATTACGT ACCGGGAAGT TTTGGAGAGC GGAAGTGGCG GCTCCCGGGA GCTGGGGACG	1620
	TCGGATTCCA GCCTCCAGGA CATCACGGAG GCGGCGGGCC ACTGCCCCGT GCATTTGTTC	1680
15	AAGGACCACG TAGACAATGA CAAGGAGAAA CTGAAAGAAT TCGGCACCGC GAGAGTGGCA	1740
	GAAGGTAAGT TCCTTTGCGC GCCGGCTCCA GGGGGGCCCT CCTGGGGTTC GCGCCTCCT	1800
	CGCCACGGAG TCGGCCCCGC GCGCCCCCTG CTGTGCACAT TTGCAGCTCC CGTCTCGCCA	1860
20	GGGTAAGGCC CGGGCCGTCA GGCTTTGCGT AAGAAAGGAA GGAAGGCAGG AGTGGACCCG	1920
	ACCGGAGACG CGGGTGGTGG GTAGCGGGT GCGGGGGGAC CCAGGGAGGG TCGCAGCGGG	1980
	GGCCGCGCGC GTGGGCACCG ACACGGGAAG GTCCCCGGCT GGGGTGGATC CGGGTGGCTG	2040
25	TGCCTGAAGC CGTAGGGCCT GAGATGTCTT TTTCAATTTT TTTTCTTTT CTTTCCTTTT	2100
	TTTGTTTGTT TGTTTGTTT TTTGAGACAG AGTCTCGCTC TGTCCCCAG GCTGGAGTGC	2160
	AGTGGTGCAG TCTCGGCTCA CTGCAACCTC CGCCTCCTGG GTTCAAGCGA TTCTCCTGCC	2220
30	TCAGCCTCCC CAGTAGCTGG GATTACAGGC ATGCACCACC ACGCCTGGCT AATTTTGTG	2280
	CTTTTAGTAA AGACGGGGAT TCACCATGTT GGCCAGGCTG GTCTCGAACT CCTGACCTCA	2340
	GGTGATCCAC CCGCCTCGGC CTCCCAAAGT GCTGGGATGA CAGGCGTGAG GCACCGCGCC	2400
35	CGGCCTGGGT CCGTACGGCT TAGGATGTGT GTTCTGTCT CTGCCTGTCT GCCTTGATT	2460
	TACGGTCACC CAGACGCACA GAGGAGCCGT CTCCACGCGC CTTCCAGCG CTCAGCGCCT	2520
	GCCGGGCCCC CGGAGATCAC GGAAGACTC GAGGCTGCGT GGTAGGAGAC GGAAGGCCCC	2580
40	CGGGTCAGCT CGGTCTGTT TCCTTTAAGG AACCCTTCAT TATTATTCA TTGTTTCCT	2640
	TTGAACGTCG AGGCTTGATC TTGGCGAAAG CTGTTGGGTC CATAAAAACC ACTCCCGTGA	2700
	GCGGAGGTGG CCGGATCTG GATGGGGCGC GAGGGGCCCC GGGGAAGCTG GCGGCTTCGC	2760
45	GGGCGCGTCC TAAGTCAAGG TTGTCAAGC GCAGCCGTT GTGCGCGGCC CGGGGAGCT	2820
	CCCCCTGGC CCTTCCTCCT GAGACCTCAG TGGTGGGTCG TCCCGTGGTG GAAATCGGG	2880
	AGTAAGAGGC TCAGAGAGAG GGGCTGGCCC CGGGGATCTC TGTGCACACA CGACAACTGG	2940
50	GCGGCATACA TCTTAAGAAT AAAATGGGCT GGCTGTGTCT GGGCACAGCT GGAGACGGCT	3000
	ATGGACGCTT GTTATGTTTT CATTACAAAG ACGCAGAGAA TCTAGCCTCG GCTTTTGCTG	3060
	ATTCGCAGAG TTGAGGTGCG AGGTGAATG CCCCAAAGGT AATCTTCCT AAGACTCTGG	3120
	GGCTACCTGC TCTCCGGGGC CCTGCATTTG GGGTGTGGAG TGGCCCCGGG AAATAGCCCT	3180
55	TGTATTTCGT GGAGGCACCA GGCAGCTTCC CAAGGCCCTG ACTTTGTCTG AGCAGAAAGC	3240

EP 1 260 228 A2

	TGTGGCTACG	GTTTACAAAG	CAGTCCCCGG	TTTCTGACCG	TCTAAGAGGC	AGGAGCCCAG	3300
5	CCTGCCTTTG	ACAGTGAGAG	GAGTTCCTCC	CTACACACTG	CTGCGGGCAC	CCGGCACTGT	3360
	AATTCATACA	CAGAGAGTTG	GCCTTCCTGG	ACGCAAGGCT	GGGAGCCGCT	TGAGGGCCTG	3420
	CGTGTAATTT	AAGAGGGTTC	GCAGCGCCCG	GCGGCCGCTT	CTGTGGGGTT	GCTTTTGGT	3480
10	TGTCCTTCGC	AGACACCGTT	TTGCTCCTCT	GAACTCTCTC	TTCTCCCCCT	GGCCGTGGAC	3540
	CCGGGAGAGC	AAAGTGTCTT	CCAGACCTTT	TGAAAGTGAG	AGGAAAATAA	AGACCAGGCC	3600
	AAAGACCCAG	GGCCACAGGA	GAGGAGACAG	AGAGTCCCCG	TTACATTTTC	CCCTTGCGTG	3660
15	GGTGCAGAAA	GACCCCCGGG	CCAGGACTGC	CACCCAGGCT	ACTATTTATT	CATCAGATCC	3720
	AAGTTAAATC	GAGGTGGAG	GGCAGGGGAG	AGTCTGAGGT	TACCGTGGA	GCCTGGAGTT	3780
	TTTGGGAACA	GCGTGTCCCC	GCCGAGCCTG	GGAGCCCGTG	GGTTCTGCAA	AGCCTGCGGG	3840
20	TGTTTGAGGA	CTTTGAAGAC	CAGTTTGTCA	GTTGGGCTCA	ATTCTCGGGG	TTCAGACTTA	3900
	GAGAAATGAA	GGAGGGAGAG	CTGGGTCGT	CTCCAGGAAA	CGATTCACTT	GGGGGGAAGG	3960
	AATGGAGTGT	TCTTGCAGGC	ACATGTCTGT	TAGGAGGTGA	AACAGAATGT	GAAATCCACG	4020
25	TTGGAGTAAG	CGTCCAGCGC	TGAATGTAGC	TCGGGGTGGG	GTGGGAGGGC	CCTGGTGTGG	4080
	ATCGTGGAAG	GAAGAAAGAC	AGAACAGGGT	GCTAGTATTT	ACCCCGTTCC	CTGTAGACAC	4140
	CCTGGATTTG	TCAGCTTTGC	AAGCTTCTTG	GTTGCAGCGG	CCTTGCCCTG	GCCCCTTTGA	4200
30	GACTGTTTCC	AGACTAAACT	TCCAAATGTC	AGCCCCCTAC	CCTTGACAGC	AAGGGACATC	4260
	TCATTAGGGC	ATCGCGTGCT	TCTCATCTGT	GCTCAGCAGG	CCCGAGATAG	GAACAGAGGG	4320
	GCGTTGGAGA	TGCCACTTCC	ACCAGCCCTG	GGTTGAAGGG	GAGCGAGGGA	GACACCTTTT	4380
	ACTTAAACCC	CTGAGCTTGG	TCAGAGAGGC	TGAATGTC TA	AAATGAGGAA	GAAAAGGTTT	4440
35	TTACCTGGA	AACGCTTGAG	GGCTGAGTCT	TCTGCCCTTC	TGACTCCCCC	AGCAAATACA	4500
	GACAGGTCAC	CACTACTTGG	AGATGAGAAA	GTGCCATTTT	TGGCACACTC	TGGTGGGGTA	4560
	GGTGCCCCGAC	CGCGTGTGAA	AAAGTGGGAA	GGAGAGATTT	CTGCGCACGC	GGTTCAGCCC	4620
40	CCAGGCGCGG	TGGCGCATTC	AGGTACTCAG	ACGCGGTTCT	GCTGTCTTGC	TGAGAAACAG	4680
	GCTTCGGGTA	GGGGCTCCTA	GCTCCGCCAG	ATCGCGGAGG	GACCCCCAGC	CCTCCTGCGC	4740
	TGCAGCGGTG	GGGATAGCGT	CTCTCCGTAG	GCCTAGAATC	TGCAACCCGC	CCCGGGTCCT	4800
45	CCCCGTGTCC	TTCCCGGGCG	TCCCGCCGGG	GATCCACAG	TTGGCAGCTC	TTCTCAAAT	4860
	TCTTTCCCTT	AAAAATAGGA	TTTGACACCC	CACTCTCCTT	AAAAAAAAAA	AATAAGAAAA	4920
	AAAGGTTAGG	TTATGTCAAC	AGAGGTGAAG	TGGATAATTG	AGGAAACGAT	TCTGAGATGA	4980
50	GGCCAAGAAA	ACAACGCTCG	TGCAAAGCCC	AGGTTTTTGG	GAAAGCAGCG	AGTATCCTCC	5040
	TCGGCTTTTG	CGTTATGGAC	CCCACGCAGT	TTTTGCGTCA	AAGCGCATTG	GTTTTCGAGG	5100
	GCCCCCTTTC	CACCGCGGGA	TGCACGAAGG	GGTTCGCCAC	GTTGCGCAAA	ACCTCCCCGG	5160
55	CCTCAGCCCT	GTGCCCTCCG	CTCCCCACGC	AGGGATTTAT	GAATGCAAAG	AGAAGCGCGA	5220

EP 1 260 228 A2

	GGACGTGAAG TCGGAGGACG AGGACGGGCA GACCAAGCTG AAACAGAGGC GCAGCCGCAC	5280
	CAACTTCACG CTGGAGCAGC TGAACGAGCT CGAGCGACTC TTCGACGAGA CCCATTACCC	5340
5	CGACGCCTTC ATGCGCGAGG AGCTCAGCCA GCGCCTGGGG CTCTCCGAGG CGCGCGTGCA	5400
	GGTAGGAACC CGGGGGCGGG GCGGGGGGGC CCGGAGCCAT CGCCTGGTCC TCGGGAGCGC	5460
	ACAGCACGCG TACAGCCACC TGCGCCCGGG CCGCCGCCGT CCCCTTCCCG GAGCGCGGGG	5520
10	AGGTTGGGTG AGGGACGGGC TGGGGTTTCTT GGACTTTTGG AGACGCCTGA GGCCTGTAGG	5580
	ATGGGTTTCAT TGCCTTTGTT TTTCACCAAC AGCAAACAAA TATATATACA TATATATTAT	5640
	ACAAATAACA AATAAATATA TATGTTATAC AGATGGGTAT ATTGTATATA TTATAGATAT	5700
15	TTGTTCTGTC TTGGTGCAAA GACACCCGGT GAACCCATAT ATTGGCTCCT GACTGCCTTC	5760
	GGTTCCCCTG GGATTGGTTA TAGGGGCAAC ACATGCAAAC AAAACTTTCC CTGGATTATA	5820
	CTTAGGAGAC GAAGCTACAG ATGCGTTTGA TCCAGAGTGT TTTACAAGAT TTTTCATTTA	5880
20	AAAAAAATG TGTCTTTTGG CCCCTGATTC CCCTCCGTCT TCCCGTGTGG CTGCATTGAA	5940
	AAGGTTTCCT TAGGATGAAA GGAGAGGGGT GTCCTCTGTC CCTAGGTGGA GAGAAACAGG	6000
	GTCTTCTCTT TCCTCCGTTT TTTCACCTAC CGTTTCTATC TCCCTCCTCC CCTCTCCAGC	6060
25	CCTGTCTCTT GCTACAAACC ACCCCCTCCT CCCTCCGGCT GTGGGGAGCG CAGGAGCACG	6120
	TTGGGCATCT GGATGAGCGG AGACTATTAG CGGGGCACGG GGGCTCCCCG AGGAGCGCGC	6180
	GAATTCACGC TGCCCCATGA GACCAGGCAC CGGGGGCGGG AGGGGCCTTG GGTGTCCGCA	6240
30	GAGGGACGGG CGGGCAGAGC CTTCTCCGC ATTCTAAACA TTCACTTAAA GGTATGAGTT	6300
	TATTTTCAGG GTGCTGCTGG GAGAGCCTCC AAATGGCTTC TTCCAGCCCC TGCCTGACAG	6360
	TTCAGCTCCC CTGGAAGGTC AACTCCTCTA GTCCTTTCTC CTGGTTCTGG GCAGGACAGA	6420
35	AGTGGGGGGA GGGAGAGAGA GAGAGAGAGA GAGAGAGACG GTCAGGATCC CCGGACCTG	6480
	GGGAACCCGT CAAAAATAA TGAAATTAAG ATTGCCGACC AGAGAGAGAA CCGTGACAAA	6540
	GCAAACGGCG TTCAAAGCAA AGAGACGAAC TGAAAGCCCG TTCCCGTAGG ACTGGTTATG	6600
40	AGGTCAACAC ATTCAAACAC AGCTTGCTCT GGATTTTGCT GAGCAGAGGA AGATACAGAT	6660
	GCATTTGATC CAAAGTGTGT TACATCTTTC ATTATATGTG TGTCTATATA TATAAACATA	6720
	TATAAATATA TAAACATACA TAAATGTATG TAAATATATA TAATCTATAT ACATATATAA	6780
45	ATATATAAAC ACATATATAA TATATAAATC TATAAACATA TATAATATAT AAACATAAAT	6840
	ATATAAACAT ATATAATATA TAAATATATT AACATATATA AAATATGTAT AAATATATAT	6900
	AAACATATAA ACATATATAA ATATATAAAC ATATAAATAT ATAAACATAT ATAAATATAT	6960
50	ACAAACATAT TGTATATATA TAAATATATA TAAAAACATA TATATACATA TAAAAATATA	7020
	TATAAACATA TATACATATA AAGAAATATA TATAAACATA TATACATATA AATATACATA	7080
	TATAAACATA TATATACATA AAATATATAT AACATATAT ACATATAAAA ATATATATAT	7140
	ATTAACATAT ATATACATAT AAAAAATAT ATATTAACAT ATATATACAT ATAAAAATAT	7200
55	ATATATATTT TTGGCCCTCG ATTCCCTTCG GTTCTGTGG GATGGGTGAT TGAGTCAACA	7260

EP 1 260 228 A2

	CATTCAAACA	CAACTTTTCC	ATCGATGTTG	CTTAGGAGAT	GAGGATACAG	ATGCGTTTGA	7320
5	TGGAGAGGGT	TTTACAAGCT	CTTTCATTTA	AATATATATA	TATATATATA	TATATTTTTT	7380
	GGCTCCTGAT	TCTCTTCCGT	CTTCCCATGT	GGCTGCATTT	TAAAAGGCTT	CCCTAAGATC	7440
	GTTACGATTA	AATCAACCCT	CCCCAGGCAT	CTTTACCGAG	GGCTGTGGTC	CCCAAAGCGA	7500
10	TACAGCCCAG	GAGGGAGAGA	GGCTTTGGTG	ACTTGGAGGA	AGGACTGTGT	CCCTCCTTAG	7560
	GGCGTCTGTG	GCCTCAGTGA	GGGAAGGAAG	CTGCATCAGA	CAGGGGTTC	CTCGCTGTCC	7620
	ACCCCTCTGG	CAGAAGATGG	ATTGGGCTGC	CCCGTATAAA	TTAATGAAAA	GATTAAAGTT	7680
15	TCGCTAAAGG	GGACATCGAG	TTTATGTGTC	ATCTCCTGGT	GTCTGTGTGC	CTGGGATCTG	7740
	CAATATATCC	CAGCCCTTGA	TGTACTGTTT	CTATAAAAAT	AAATTACTTG	TAATTTAATT	7800
	CCACACTATT	TCTTTCCGTA	GTCTATTACC	GACGAGAGCA	CGTTAGTTCA	GCTGCGGAAA	7860
20	ATTGGTTGTG	GGGTGTGTGC	GGACCCCGAG	AACGCCCTAA	AATAAGACA	AATCGGGGAC	7920
	AAGCTGGGGG	TTATCGATTG	CAGGGGTGCG	ATGAAAATTT	AACGACGGTA	AATAATAATA	7980
	AAAACAAACA	TGGGAATGCA	ATAAAAGACA	TAATTCTCCA	TCGCCGCGGG	GGGAAAGGAT	8040
25	CCTATAGTAA	AGGCGAGTGC	GCTTTGAGGG	GTCATAAAAA	TCAATTAGTT	CCAACACCCA	8100
	CGTCCCGCGT	TGAGGGGACG	GGGACGAGCA	GGGACAGAAA	AAGAAACCAT	ATTTGAATCC	8160
	CATCTCTCTG	TGAATTCTTG	GGTCACATGC	GTCTCAGTAC	AGCCCGTCCC	GTGCTGTGAC	8220
30	CGGATAGAGT	TTCAATTTAC	TGTGGAAATT	TGCTGTAAAT	AAATTGAGCA	TCCGATAGAA	8280
	GCTGTTGCTG	ATTAACCTTT	TATTTTTAGC	GTGGCCCTGC	AAAGTCGTAT	CACCCAGCTG	8340
	TCAGGCTTCT	AATCGAAAGT	TATGAGACCA	CGGTGAGGGG	CAGGCGGTAA	TTTAATTACA	8400
35	ACAAATATCT	TTGGGTTTAT	GGCGCAGAGC	TAAATTAAAT	GTCATTATTC	ACTGTCTGTA	8460
	ATGGAAATCA	AAAGGAAATC	GCATTACGGC	ATTTGGGAAA	GAAAGCGGGG	AGTGCTCTTT	8520
	AATGAAGAAA	TAAGTGTCTT	AAGCAGTGTC	ACACACTTCA	CTTACCATAT	TCGGGCCTAA	8580
	TTGGAATGGA	TCGTGAATCA	CTCCAAGACT	GATTTATTAG	CGCTTCACGC	AGCGGCTAAT	8640
40	TCATCACTTG	TATTCCTCAT	CATTTTTTTT	TTTCTCTCTG	CCGTGTTGAA	GGGAGAGTGA	8700
	ATGAGGCTTT	CCACGTTTCA	GGAGGATTTT	CTTTTTTGAA	AAATGCCCTT	CCAGAGGCTT	8760
	TTGGGTGGCT	GGCTTGCTTT	CTGGGCCCTG	GAGGAGACAG	GCGGAGAGTC	CAGGTGGGCA	8820
45	TGGAGAGGCA	CAGTGGCAGG	TCACCTGGAT	GGTCAGTGGA	GGTGGAGGTC	TGAAGGCGCC	8880
	AGCTTTGGAA	ATTATTGGTG	AATTTTCGATG	TCAGCACCCAG	GCAGGGCCCT	TTTTGGCGGG	8940
	GGTGTGAGGG	AGGATGACTT	TGCTGGGAAA	CAGGATCAGG	TTCTCCAGGC	GCACTGCAGC	9000
50	CCGGTAGGAC	CCACTTTGGA	AATGAAAAGC	CAGTTCCGAA	AGCTGGGCTG	GAAGCTTCCG	9060
	TGTTGGGTTT	AAGAGCAAGT	TCACGTTGCG	CTGTGTAGAC	TCCTGGCTGC	TCCCAAATCT	9120
	TGAGGGTTTT	CTGAGGTTCC	CTTCATAGGG	GCACCGGCCC	TGGGCCATGC	ACAGTGCCTA	9180
55	AGGGTGGCTG	TGGGCCGAGG	GACCCAGCAC	GTGTTTTGCC	CACAACAGCC	GGAGTGACTG	9240

EP 1 260 228 A2

	GTTCACTCAC CGCCTTGGCG GAGGACGCCT GTTCTCTGGA CGAATCATTT CTCTTGGGTG	9300
	GTGACTGCCT TGTGGGTCAA GGTGCAGGTT TTCTGCCACA GAAAACCTGT TAGGAGGAAT	9360
5	TAAGCGACTA AGACTGTCAG GGAGGTGGTG GTGGGGGAGA GGAGGGGGTG GTGTCCAGAT	9420
	TACCAGGCAT AGGCTAAACT GCCTGCACTC TCCAGCTGGT CTGTCTGTGG AGGAGGGGAT	9480
	TGTCAATACT GGGAGAGCAG AGGAGGCTCG TAGGAGGTGA GAGGGGGTGG AATTTGCATG	9540
10	CAAATCTTCA CATGAGGCCT GTGTGAATTT CTCCAGCCTC CTGAGGGTCC CCTGCGCTAT	9600
	TGCACTCAAC TTCTTGATAG TTTACCCCAA GACTCAGAAG TCCTTAGAGG GGCAGAATGC	9660
	CCCCACCACA AAGCCTGCTA TCCTTGGGCG TCCTCAGGAC CCTTGGTCAT GAATGGGACC	9720
15	CTTTCATGTA TGGGGACCCT TGGTAATATG AATGGGACGC CTTCAGCTCC CCAGGGCTTC	9780
	CGAGGAGGCC GAGAAGGGCA AAGACACTTC CGAGGAGGCC GAGAAGGGCA AAGACATTTT	9840
	CTGGGCTTGG TGTGTCAAGA GCTAGATTGG AGAAGGGGCT GGATTTGGA CTCTTTAGCC	9900
20	ATCAGCTCAC CCTCTCCGTT TGTGGCTAAA GTCTGAAGGT GGAAACTTCG GTTCTCCTAC	9960
	AGGGTCTACA GGAGTTGGGG GCGGGGCGC CCACACAGAA CGCTGGAAAG TTCGACAGTC	10020
	CACTTCCACT GGCTCGGAAC TCACTTTTTC ACCTTAAGTT CATCAGCGGT AACGCATAGG	10080
25	TCTCACTTAG GCAGGGCACG GATGATTTAA CAATTTCTAC TTCTAGGTCA GGTGCGGTGG	10140
	CTCACACCTC TAATCCAGC ACTTTGGGAG GCCCAGGAGG GTGGATCGCT TGAGGTCAGG	10200
	AGTTTGAGAC CAGCCTGGCC AACATGGTGA AACCCGCTCT CTACTAAAAT ACGAAAATTA	10260
30	GCCAGGCATG GTGGTGAGCA CCTGTAATTC CAGCTACTCG GGAGGCTGAG GCAGGAGAAT	10320
	CGCTTGAACC TGGGAGGTGG ACGTTGCAGT GAGGTGAGAT CACACCACTG CACTCCAGCC	10380
	TGGATGAGAG AGCAAGACTC TGTCTCAAAA ACAAAATAAA ACAAAAACAA AAAAAAATC	10440
35	AAAAAAGAAA ACCCAATTTT CAGTTCTAGG CCAGGTGCAG TGGCTCACGC CTGTCATCCC	10500
	AGCACTTTGG GAGGCCAGG AGGGTGGATC GCTTGAGGTC AGGAGTTCGA GACCAGCCTG	10560
	GCCAACATGG TGAAACCCCA TCTCTACTAA AAATACAAAC GTTAGCTGGG TGTGGTGGTG	10620
40	TGCGCCTGTA ATCCCAGCTA CTCGGGAAGC TGAGGCTGGA GAATTGCTTG AATCTGGGAG	10680
	GTGGAGGTTG CAGGGAGGCG AGATAGTGCC ACTGCAGTCC AGCCTGGACC AGAGAGCAAG	10740
	ACTCCGTCTC AAAAACAAAA GAAAGCAAAA ACAAAAACA AGAGACCAGC CTGGCCAACA	10800
45	TGGTGAAACC GCGTCTCTAC TAAAAACAA AATTAGCCGG GCATGGTGGT GGGCACCTGT	10860
	AGTCCCAGCT ACTCGGGAGG CTGAGGCAGG AGAATGGCTT GAACCTGGGA GGTGGAGCTT	10920
	GCAGTGAGCC GAGATAGTGC CACTGCACTC CAGCCTGGGC GACAGAGCGA GACTTGATTT	10980
50	CAGAACCACC ACCACCACAA CAAAACAAAA CAAAAAATCC AAAAAACCC CAATTTCCAG	11040
	TACTAGGTAG TCAGTGATGC AGGGCTGGAG ACAGAGGGGC GGTAAGTGTC TGGGCGCCCA	11100
	CCATCAGTCA CCTCCCAGCT CCCAGAGGTG CAAAGTGCTT GGTTCAGCCT CATGGGAAGG	11160
55	ATGCTCCCTG GGGAGGCTGG GCTGGGTTC CAGGCTCTT CACATCTCTC TCTGCTTCTC	11220
	CCCAAGGTTT GGTTCAGAA CCGGAGAGCC AAGTGCCGCA AACAAGAGAA TCAGATGCAT	11280

EP 1 260 228 A2

	AAAGGTGGGT	GTCGGGACTG	GGGGGACCTG	AAGCTGGGGG	ATCCTGCTCC	AGGAGGGATG	11340
5	GGGTCGACGA	GGTGCTGGCT	ACACCCAGGA	CCACCACACT	GACACCTGCT	CCCTTTGGAC	11400
	ACAGGCGTCA	TCTTGGGCAC	AGCCAACCAC	CTAGACGCCT	GCCGAGTGGC	ACCCTACGTC	11460
	AACATGGGAG	CCTTACGGAT	GCCTTTCCAA	CAGGTAGCTC	ACTTTTCTCT	CCTCTGAAGA	11520
10	TCCCTAGGGA	CCTGCTGCTC	CCTTCCCCTT	TCCCCTATTT	GCTGCCGCAT	CCTGACACTC	11580
	CTAGTCCCTC	CCTGCCCCCTG	CAGACTTCTC	AGCTGGCCCT	TAGAAAAAAA	GCCTCTTTTC	11640
	CGAGGAGGCA	TTTACAGGCA	CCTTGGCACC	TATGAAATCA	GGCTGGGCCA	GGCGGGGTGG	11700
15	CTCACACCTG	TCATCCCAGC	ACTTTGGGAG	GCTGAGGAGG	GTGCATCACC	TGAGATCAGG	11760
	AGTTCAAGAC	CAGCCTGGCC	AACTTAACGA	AACCCCGTCT	ATTAAAAATA	CAAAATGGGT	11820
	GTGGTGGCTC	ACGCCTGTCA	TCCCAGCACT	TTGGGAGGCC	GAGGCAGGTG	GATCACCTGA	11880
20	GGTCAGGAAT	TCGAGACCAG	CCTGACCAAC	ATGCTGAAAC	CCCGTCTCTA	CTGAAAACAC	11940
	AAAGCTTAGC	CGGGCGTGGT	GGTGACACACC	TGTGATCCCA	GGTACTTGGG	AGGGAGAATC	12000
	ACTTGAACCT	GGGAGGTGGA	GGTTGCCGTG	AGCCAATATC	GCGCCACTGC	ACTCCACTCT	12060
25	GGGTGACAGA	GTGAGACTCC	AAGACTCCAT	CTCAAAAAAA	AAAAAAAAAA	TCAGGCTGTA	12120
	AAAATCCACT	TTTGGGAAGG	TGAACACACA	CAAGCCCAA	CAGAAATCTG	ACAAAAACCA	12180
	GAGGGGTGAA	AAGTCCACAC	AGTCAGGCAC	CCCCACCTGG	CTTGCTGCCT	GGTTAAGAAG	12240
30	GGCGCAGATG	CCTGTGCCTG	GATACCAGAG	ATGGGACAGA	CACCCATTCC	CTTTTCATCA	12300
	CCACCCCCGA	GTGCCCAGAG	GCCTGGGGCG	TC TGCTGGC	CCCTGGCCCC	TGGCTTGGGC	12360
	TCTGCACCTC	TGAAGTGGAG	ACACCCTACT	CAGCTCCCCA	CTTACTTTGG	AGTGAGCAGC	12420
35	GCTTGGGTGC	CCAGCGTGGG	TTTGGGGCTT	CCAGGGAGTC	GGGGTTGGGT	CGCGGAGCCC	12480
	AAGCTTCCCA	AGGGCGCCCC	CGCCCTGCCC	TGGCTTAGTG	GTGGGGATGG	GATGGGGGGA	12540
	AACGGGGAGC	TGCGTGGAAG	GAGGTGAAGG	GTACAGGAG	GAGAGAGCGC	AGCGCCACG	12600
40	TGCGCCCTGC	CTGAACGCGC	AGCGCAGCGC	CCGGCTGCGG	TGCCCCCTGC	CCCTTCGGTC	12660
	CCTAATTG	GGATCGGGAG	TGCATGCGCG	GGCGGAACGG	GCTTGGGGGG	GGGGCTCTGG	12720
	CAGGGCGGAC	GCGTGGCCTC	CCTTCTTCAC	CGTTTTATTC	CAAGGGGACA	GGCTGGGGAT	12780
45	TGTATTGGG	CGCGTGTG	GCTGAGGGTG	CAGGGACTTG	GGGGGTGGCG	GTGGGGAGCG	12840
	CGGAAGGTAT	AAACGTATAA	ATCATAAGTA	AACAACCTAG	AAATGGACCC	CGAGCGCTGG	12900
	TCGCCGCTAG	CTCTCCAGCT	CTCCCTGGCC	CAGGCCCGAA	GGAGAGGGGT	CCGCATCCCT	12960
50	CCGCGGTTCT	CCTCTCCTGG	GTACCTGGCC	TTGAGGTGGG	GGAACGAGCC	TACTTCTTGT	13020
	ACCGTCTTTT	GCCGACGGCG	GGACCCAGTG	AAATTAGGCC	GTTGGAGCCC	GCAGGCCTGC	13080
	CTGGCTTTGC	GCACCGGAGT	CTTGGGGACC	TGGTGTCCCC	GGGAAAAACT	TGGGGACCTG	13140
	GTATCCCCGG	GAGAGGCTTG	GGGACCTGGT	GTCCCGGGAG	AGGCTTGGGT	ACCTGGTTTC	13200
55	TCTGGAAGAG	GCTTGGACAC	CTGGTGTCTT	GGGAGGGCCT	TTGGGACCTG	GTGTCTGGG	13260

EP 1 260 228 A2

	AGAGGCTTGG	AGATCTGTTG	TCCTGGGAGA	GGCTTGGGGA	CCTGGTGTCC	CTGGAGAGGC	13320
	TTGGGGACCT	GGTGACCTTG	GAGAGGCTTG	GAGACCTGGT	GTTCCTGGGAG	AGGCTTGGGG	13380
5	ACCTGGTGT	CTGGGAGAGG	CTTGGGGACC	TGGTGTCTCT	GGAAGAGGCT	TGGACACCTG	13440
	GTGACCCGGG	AGGGCCTTGG	GGATCTGGTG	TCCCGGGAGA	GCCTTGGGGA	CCTGGTGTCC	13500
	TGGGAGAGGC	TTGGGGACCT	GGTGACCTTG	GAGAGGCTTG	GGGACCTGGT	GTCTTGAGAG	13560
10	AGCCTTGGGG	ATCTGGGTGC	CCAGGAGAGG	CTTGGGGACC	TGGTGTCTCT	GGAAGAGGCT	13620
	TGGACACCTG	GTGTCTGGG	GAGAGGCTTG	GGGACCTGGT	GTCTTGGGAG	AGGCTTGGGG	13680
	ACCTGGTGTG	CTGGGAGAGG	CTTGGAGATC	TGGTGAGCCG	GGAGAGGCTT	GGGGACCTGG	13740
15	TGTCCCGGGA	GAGGCTTGGG	GACTTGGTGT	CCCGGGAGAG	GCTTGAACAC	CTGGTGTCCC	13800
	AGGAGAGGCT	TGGGGACCTG	GTGACCTTGG	AGAGGCCTGG	GGACCTGGTG	ACCCGGGAGA	13860
	GCCTTGGGGA	CCTGGTGTCC	TGGGGAGAGC	CTTGGGGACC	TGGTGACCTT	GGAGAGGCTT	13920
20	GGGGACCTGG	TGTCCTGGGA	GTGCCCTTGG	GACCTAGTGA	CCCGGGAGAG	GCTTGGGGAC	13980
	CTGGTGTCCC	GGGAGAGGCT	TGGGGACCTG	GTGTCTGGG	AGAGCCTTGG	GGATCTGGTG	14040
	TCCTGGGGAG	AGGCTGGGGG	ACCTGGTGTG	TCGGGAGAGA	GCCTTGGGGA	CCTGGTGACC	14100
25	CGGGAGAGGC	TTGGACACCT	GGTGTCCCGG	GAGAGGCTTG	GGGACCTGGT	GACCCGGGAG	14160
	AGCCTTGGGG	ACCTGGTGTG	CTGGGGAGAG	GCTGGGGGAC	CTGGTGTCTC	GGGAGAGAGC	14220
	CTTGGGGACC	TGGTGACCCG	GGAGAGGCTT	GGACACCTGG	TGTCCCGGGA	GAGGCTTGGG	14280
30	AGCCTGGTGT	CCCGGGAGAG	CCTTGGGGAC	CAGGTGACCT	TGGAGAGGCT	TGGGGACCTG	14340
	GTGATCTTGG	AGAGGCTTGG	GGACCTGGTG	TCTCGGGAGA	GGTTACGGGG	GCTGGTTGGG	14400
	GGAGAGAACG	TTGTGAGCCA	AAGTCCCTGA	ATCCCTGCGA	AAAGAGCGCA	TCGGGAGCTC	14460
35	CCCCTGAGGG	CGTTCCATTT	GTGGACCCCC	CTCCCATGCG	CTTTGCAGGG	AGCTGTTCGG	14520
	ATTCCCTTGG	CCCGGCTCCC	GCGGATGCAT	CCAGTGGCAG	CGCCAATTCT	GGGCCAGGGG	14580
	GAAGGAGGAA	AGGCGGGTGT	GGGGTGGTCT	CCACGGCTGG	AGAAGGGGCG	ACGCTCCCTA	14640
40	GGGGAGAAGA	GGCACGTTGG	GGGTTCCTGG	GGGCGCGGGG	CGGAGCAGGC	CCCCAGTCC	14700
	CCATCTGCG	CCCTCACCCC	GCCGGGTCCG	CTCCCGCAGG	TCCAGGCTCA	GCTGCAGCTG	14760
	GAAGGCCGTG	CCCACGCGCA	CCCGCACCTG	CACCCGCACC	TGGCGGCGCA	CGCGCCCTAC	14820
45	CTGATGTTCC	CCCCGCCGCC	CTTCGGGCTG	CCCATCGCGT	CGCTGGCCGA	GTCCGCCTCG	14880
	GCCGCCGCGG	TGGTCGCCGC	CGCCGCCAAA	AGCAACAGCA	AGAATTCCAG	CATCGCCGAC	14940
	CTGCGGCTCA	AGGCGCGGAA	GCACGCGGAG	GCCCTGGGGC	TCTGACCCGC	CGCGCAGCCC	15000
	CCCGCGCGCC	CGGACTCCCG	GGCTCCGCGC	ACCCCGCCTG	CACCGCGCGT	CCTGCACTCA	15060
50	ACCCCGCCTG	GAGCTCCTTC	CGCGGCCACC	GTGCTCCGGG	CACCCCGGGA	GCTCCTGCAA	15120
	GAGGCCGTAG	GAGGGAGGCT	CCCGGGACCG	TCCACGCACG	ACCCAGCCAG	ACCTTCGCGG	15180
	AGATGGTGCA	GAAGGCGGAG	CGGGTGAGCG	GCCGTGCGTC	CAGCCCGGGC	CTCTCCAAGG	15240
55	CTGCCCCGTG	GTCTTGGGAC	CCTGGAGAAG	GGTAAACCCC	CGCCTGGCTG	CGTCTTCCTC	15300

EP 1 260 228 A2

	TGCTATACCC	TATGCATGCG	GTAACTACA	CACGTTTGA	AGATCCTTAG	AGTCTATTGA	15360
5	AACTGCAAAG	ATCCCGGAGC	TGGTCTCCGA	TGAAAATGCC	ATTTCTTCGT	TGCCAACGAT	15420
	TTTCTTTACT	ACCATGCTCC	TTCCCTCATC	CCGAGAGGCT	GCGGAACGGG	TGTGGATTGT	15480
	AATGTGGACT	TCGGAATCCC	AGGAGGCAGG	GGCCGGGCTC	TCCTCCACCG	CTCCCCCGGA	15540
10	GCCTCCCAGG	CAGCAATAAG	GAAATAGTTC	TCTGGCTGAG	GCTGAGGACG	TGAACCGCGG	15600
	GCTTTGGAAA	GGGAGGGGAG	GGAGACCCGA	ACCTCCACG	TTGGGACTCC	CACGTTCCGG	15660
	GGACCTGAAT	GAGGACCGAC	TTTATAACTT	TTCCAGTGTT	TGATTCCCAA	ATTGGGTCTG	15720
15	GTTTTGTTTT	GGATTGGTAT	TTTTTTTTTT	TTTTTTTTTT	GCTGTGTTAC	AGGATTTCAGA	15780
	CGCAAAAGAC	TTGCATAAGA	GACGGACGCG	TGGTTGCAAG	GTGTCATACT	GATATGCAGC	15840
	ATTAACTTTA	CTGACATGGA	GTGAAGTGCA	ATATTATAAA	TATTATAGAT	TAAAAAATAA	15900
20	ATAGCCGTGC	ACTCTTGACC	CCGTCAACGT	CCAACGTGGA	AAAGGCGTTA	CCTCTTCTCC	15960
	CAGCGCTGGC	CGCCTGGCCA	CTGAGGGCCC	TTTGCAAAAA	TCACGGGTGT	AGAGATGGCC	16020
	CTGGGCGCGC	TGGGAGTGTG	GTTGTGTTTC	TGAAGGGGAT	AAAAGAGGGC	ACGGTGGTGC	16080
	CAAGATATCA	GTTTGGTACC	TGAGCTGTTT	CTGGTTGGGA	AGCGTAAAAG	CCAGGGAGAG	16140
25	ATCCAGAGAG	TTTTCAAGTT	TTTGCAGATG	TAGGTGGTTC	CAGCTTTTCT	TTCTCCCCTA	16200
	CTCCATCTTC	TGCGTTCCCC	CAGTTCTTTT	ATTTCTTTGT	TTTTTATTTT	TGAGACAGAG	16260
	ACTTGCTTTG	TCGCCCAGGC	TGGAGTGCA	TGGCGCAATG	TCAGCTCACT	GCCACCTCCA	16320
30	CCTCCCGGGT	TCAAGCGATG	CTCCTGCCTC	AGCCTCCCGA	GTAGCTGGGA	CTACAGGCAC	16380
	CTGCCACCAC	CCCCGGCTAA	TTTTTTGTAT	TTATAGTAGA	GACGGGGTTT	CACCGTGTGT	16440
	GCCAGGCTCG	TCTCGAACTC	CTGACCTCAG	GTGATCTGCC	CGCCTCGGCC	TCCCAACGTG	16500
35	CCCCCAGTTT	TATAAACAGC	AGATAGCAAC	TTGTCGTAC	AGCTGGCATG	GGCTGGACAG	16560
	TTGCTTGAAA	TGACCTAACC	AAAAACATTC	AAGGTTCTTG	CCCCCAGATT	TCGGGAGATC	16620
	CACGTTCCAT	GTTCTGATTG	GTTTCTCTGG	AACACAGCAA	GGGGTTTGGT	GACCTCCGAG	16680
40	AAGATCCATC	TGCATGATTG	GCATTAGTTA	CCACAGCCTG	CCCAGAGAGA	AACTATCTTC	16740
	TCCCAACATT	TACTAACATC	CACCTGGTCAA	CTCTCTTATT	TCCATAACAC	ATTTGCATCT	16800
	TTCTGGATTG	AAGCTTGGTG	GTTTCTCTTC	CTAACTTCTG	ATTTAGATAC	TTCTCCCTGA	16860
45	GGTGGGGATA	AAAGAAAAAA	AAAAAACAA	TTCTTTTTTT	CTTCCGCATA	ACACTTTCTA	16920
	TCTTGTCACT	GAGCTGAACT	GTAGATCCAT	TTGGACCCGT	CTCATTTGTA	TCTTCTGATA	16980
	TTCTTTATAC	AAACCAAAAG	TCCCCCTCAA	CATTTTATAT	GTCAAAATGT	TACAACCGCT	17040
50	GTAAAATGAC	GGAGAGAGAG	AGAAAAGAAT	CCAGACATTA	ACGGTATTAG	AGAGTTTGCC	17100
	TCATTTCATC	ATTTTCTCTA	AAAGCTGGAA	ATTAAAAAAA	AAAAAGAGAG	AGAGAGGCTT	17160
	TAATAGTTAA	GCTGAAATTT	TTATCGAAAA	GAAGAATTGC	ATTTTGAATC	TTTGGGAAGT	17220
55	AGGTTCATTC	ATCAGAGTAT	GTAACCCCTT	GGAAAAGTGG	TTGGTAAGAT	ATGTACAGCC	17280

EP 1 260 228 A2

	CTAGATTTTT TTTTTTTTAA CCAAAAAGGC TGAGTAATTT TGAAAAATCG AAACATAACA	17340
	GTGTGTCATC ATTTCTCTCC AAGAAAAAGC TCAC'TCCACG TGAGTAGAAA GACATCTACC	17400
5	TGGTCCCTGT AGAATCTGAA CGTTTCTCTT TAGAGACGGA ATTTCAATCT TGTGCCCCAG	17460
	GCTGGAGTGC AGTGGCACAA TCTCGGCTCA CCGCAACCTC CGCCTCCCGG GTTCAAGCCA	17520
	TTCTCCTGCC TCAGTCTCCC GAGTAGCTGG GATTACAGGC ACCTGCCACC AGGCCTGGGT	17580
10	AACTTTCTGG TATTTTTAGT AGAGACAGGG TTTCAGCCTC CCGAGTAGCT GGGATTACAG	17640
	GCACCTGCCA CCAGGCCTGG GTAAC'TTCT GGTAT'TTTA GTAGAGACAG GGT'TTCAGCC	17700
	TCCCAGTAG CTGGGATTAC AGGCACCTGC CACCAGGCCT GGGTAAC'TT CTGGTAGTTT	17760
15	TAGTAGAGAC AGGGTTTCGG CCTCCCAGT AGCTGGGATT ACAGGCACCT GCCACCAGGC	17820
	CTGGGTAACT TTCTGGTATT TTTAGTAGAG ACAGGGTTT CCGCTCCCGA GTAGCTGGGA	17880
	TTACAGGCAC CTGCCACCAG GCCTGGGTAA CTTTCTGGTA TTTT'TAGTAC AGACAGGGTT	17940
20	TCGGCCTCCT GAGTAGCTGG GATTACAGGC ACCTGCCACC AGGCCTGGGT AACTTTCTGG	18000
	TAGTTT'TAGT AGAGACAGGG TTTCAGCCTC CCGAGTAGCT GGGATTACAG GCACCTGCCA	18060
	CCAGGCCTGG GTAATTTTTT TGCA'TTTTG GTAGAGACAG GTTTT'TGCCG TGTGCCCCG	18120
25	GCTGGTCTCA AACTCCTGAC CTCAGGTGA CCTGCCCGCT TTGTCCCTCG CAAAGTGCTG	18180
	GGATTACAGG CGTGAGCCAC CACACCTGGC CTGAATCTGA ACTTTTAAAA GGGAGT'ACT	18240
	GACTCTCAAC TGTGCGGGGA CGGTTTCACT TTGATTTAAT ATGGAAAGAG GGCCAAGTGT	18300
30	CATCCTCACA AATGGGTCCC CGAAGCAGAT CAAACGCAGA GAACTGTGAG GGTGGGACAC	18360
	GAGTGTCTGT GGACACTGGC TGCC'TTGGC TTTTCTCCTG CGAGAGAAGT TGGGTGACTT	18420
	TCTGTAGGTG GATGAGTGAT CCCTGAATGA GTGTGGGGTA CGTGTATGCT AGCTGCTTCT	18480
35	TTCTCCCTGA AACTCTCGGA TGAAGGAAG TAAGAAATTC AGCTTGGGCT GTGACCAGTT	18540
	CTCACCACCA ACGCCCTCTT CTCTCTCCCT TCTCCTTCCT TCCTTCCTTC CTTCCTTTCT	18600
	TTCTTTTCTT TTCTTTCTCT CT'TTCTTCT TTTCTTTCTT TCTGTTTCTT TCCTTTTAT	18660
40	CTTCTCTCTT TTTCTTTCT CT'TTCTTCT TTTGTTTCTT TCTTCTTTT TCTTTCTTTT	18720
	TTTTTCTTTC TTCTTTCTTT CTTCGATGAA GTCTCACTCT GTCACCCAGG CTGGAGTGCA	18780
	GTGGTGCAAT CCCAGCTCAC TGCATCCTCT ACCTCCTGGC TTCAAGAAAT TCTCCTGCCT	18840
45	CAGCCTCCCA AGTAGCTGGG ATGACAGGCA CCCACCACCA TTCCCGGATA ATTTTGTAT	18900
	TTTTTAGTAG AGACTGGGTT TCGCCATGTT GGCCAGGCTG GTCTTGAACCT CCTGACCTCA	18960
	CATGATCCAC CCGCCTCAGC CTCCCAGAGT GCTGGGATTA CGGGGTGAGG CACCGCGCCC	19020
50	GGCCTCCTCT CTCTTTTCTT GAGATGTTTA GGAAGGACTG GGCTGATGGG GACCCTCTGT	19080
	ATGTGATGTG CGTGGGTTTG GTT'CCCCGA AGGCCCTCCA GAGACACGTT TGCCTGAACA	19140
	TTCAGCATGG AAACAACATA CGTCTCTCCA CAGGAGGTGA GAAAT'TGAAT TTATGGGGTG	19200
	GGTGTACGCT GCGGATTCCT GGTGCTTTT GCTCAAAACA AGGT'TCTTTT GAAAGTCACG	19260
55	TTCTGCTTT CCCTGTGGCT TCCCGGTGAG CTCGCTCGCA GAGCAAGGAA TACCACCCAG	19320

EP 1 260 228 A2

	AGAGCAACGT	GGGCTGTGTT	CCGTTGTAAC	GCCGTTGCAG	AGAGAGGATT	TGGTGTGTGA	19380
5	GATCCGTACC	AGCTCCAGCA	CACTGATAGG	AACACGTTGC	TGGCCGAAC	GAACGATGCT	19440
	GGGTTGGGTC	CTGATTGATA	CGTATTTTCT	TCCCTCCTCT	CCCCAAAAC	TGGCCAAATA	19500
	GTCCGTGGAG	GGTTGTCACT	CGCCGCAGTT	GAGCAAAAAA	CACCTCTTCC	TTTGAGTGGC	19560
10	TGTTCTGGTG	AAATCTGTTT	CTGACATATC	CACTTTTTCT	TCTCTTTTCT	CTCTCTCTGA	19620
	CTGCGAAGCA	CCCACAGGGA	GAAGGAATTG	GATGTATCGG	ATGTTGGTAT	TAGATTTTCT	19680
	TTCTCCGTTT	GAGTCTCTGA	CTGGTGCATA	CTTGTCAAAG	GTGTGTTTCT	GGCAATTGCC	19740
15	AAGAGTTAGA	AAAATGCACC	TTCTCTGGTG	GCCGTTGGGG	TGTTGTTTCA	CAGGCAGTGG	19800
	TGACAGGGCC	CCTTGGCTGT	GGCTGTCTTC	TCCAGCGCCG	TGGATAAAGA	GACGGGACAG	19860
	ATTCTGTGCC	TCTGTACGAT	TTAGAGCGTA	ACTGACCGCG	TCCAACACCC	GTTTTTCCAC	19920
20	TTACAAAGCT	GGTGGTGCGA	CGGGCTTGGT	GTCTCCCGTA	CGGGAAGGAG	GCCTTTGGGC	19980
	CGCTCCAAAG	ACGCCCTGTC	GTAGGAATGG	CCTCTCCATC	CCGCCAAAGT	CCAGCCAGGC	20040
	CCCCGAAATG	GTCCCATTTT	CTTGAAGGCC	TGAGTTTCTG	TTCTGGTCTT	GCTGCTGTCC	20100
25	TTGGCCACGT	CAGCACGTGG	GAGCATCTGT	GGATACCGCA	GAGTCTGGGG	ACAGCTGGGC	20160
	GTTTAAACCGA	AATGAAGCCG	AGACGGGTTT	CAGGTTTGGG	TGCCAAGCTC	TGGTCAGGAT	20220
	GAAAGGGAAA	TACCAGAGTC	CTCTGTCTCT	GCCTCTGGGT	TTTATGCTGA	CCTTTCTAAC	20280
30	ATTTGTTTTT	CCCTAAGAAC	AAGCAGAAGC	CTCCAGCTCC	CTTTAGCTCC	ACAGTTTTTC	20340
	CGGGGACATA	GCGAGGATGG	CACACGGCAG	CCACTCCAC	GACACACATT	TCGGAGGCAC	20400
	TTTGCTGGAA	GCCGCTTGTC	TCCTCCAGCT	TTGGGAGGTC	TGGGGAGGAG	AGAGGCTTTC	20460
35	GGTGGACACG	TTTGACATTA	AAAAAAAAAA	AAAAAAAAAA	AAAAAACTG	GTGCCTAATT	20520
	TATTAAAGAG	AATTAGCTTA	GCGAGTATAT	GCTGATATTC	TTTGACACAC	GTGGGTAAGT	20580
	TGATGCCATT	TATAAATGTT	TTATTGAAAT	TTGATATTTA	ATGAGAAGCC	GGTTAAGGAA	20640
40	TGTAGACAAT	ATCCCGTTTC	AAAGCTATGA	AATGTGCTAT	TTATTGAAAG	GGGATGTGGC	20700
	TTACAGAGTT	CAGCCCATTG	TACGTGCAGG	TCCCGTGGGA	AGGAGGCAAA	AGCCCCTGCT	20760
	TCTTACTTTG	TGATGTATGT	GCATTTGTTA	TTTATTTTTT	TTTCCTTGGT	CGGACGTTCA	20820
45	TAAATATGTA	CTATTTTAAT	TATGTCGAGT	GTAAATTTGA	CATCGCGTTG	CATTTATTTT	20880
	TATATTTCTG	AAAACGTGTT	CTTTTCTTTT	TTCCCTCCCC	CATTGACGAC	ATAGCGGCCC	20940
	CCGCGTCCGG	GTTACAAATA	CATCTACAGA	TATTTTCAGG	GATTGCTTCA	GATGAAAACA	21000
50	AATCACACAC	CGTTTCCCAA	ACCAACAGTC	TTACATTTTC	TATCCCTCTG	TTATTGTTCG	21060
	CAGGCGGTGA	GGGGTAGAAA	AAAAACAAC	AAACAACAG	AAAAAAAAC	CAAAAAAAC	21120
	CACCCGTAGT	TTCTCTGGTG	ACGCCCTCAT	TCTCCTAACG	TTCAATAATC	TCAATGTTGA	21180
	GTTGCAGCAA	CAGACTGTAT	TTTTGTGACG	CCCCGTAGTA	TGAATGTACA	TCTTGTAATA	21240
55	CTGAGATATA	AATAAACTTA	TAAATATTTG	TATTCAAGTG	TTAAAAAAA	AAAAATTTCT	21300

EP 1 260 228 A2

	AACCTCTCCC	CTGAGGACAG	GCTTATTGGA	AAAAAAAAAA	AAAAAAAAAA	ATCCTGAGTC	21360
	GGCCGTGGCT	GAACACAGAG	TGTTGTTCTG	CTCCGTGCAT	TTCCAGGGTG	GGTACCCAGT	21420
5	GTTGCCCCC	AGCCTTAGAT	CGGGAGGTAC	CATTGACTTT	TGCTTGATATC	CCATCCCCTT	21480
	CCTTTACTGA	AACCTACCTC	CCCCTTCTC	AGCCAACGTC	CCCCCAGAAG	GTGGCAAAAA	21540
	AAACAGAGGA	AAAAGCCCTG	ATTTGAATCA	AGTCAGAGCT	GCTAATTCTC	CACTTTCTTT	21600
10	AATTAATTAA	TTTATTTTTT	TTTTTGAGAC	TGAGTCTCGC	TCTGTGCGCC	AGGCCGGAGG	21660
	AGTGCAGGGG	CGGATCTCG	GCTCACC GCG	ACCTCCGCCT	CCCGGGTTCA	AGCGACTCTC	21720
	CTGCCTCAGC	CTCCCGAGTA	GCTGGGATGA	CAGTCACCTG	CACCACCGCG	CCCGGCTCAT	21780
15	TTTGTATTT	TTAGTAGCAA	TGGGGTTTCA	CCGTGTTGGT	CAGGCTGGTC	TCGAACTCCT	21840
	GACCTCGTGA	TCCACCCGCG	TCTGGGCCCC	GCCGGTGATG	TGTGTGCTTT	TAACTTTTAT	21900
	TTTGTTCAG	TTTTCGACAG	TGGCACGGAT	TTTCCAGCAC	GGTCTTGCAA	GGATGATTGA	21960
20	GTCATTTTTG	AGACAAAAA	TATAATAATA	ATAAATGGAA	AAAGAAATCG	ACTTTTAAAA	22020
	ATGACAAATT	TTTTTTTTTT	TTTTTTGCAT	AGATTTTTCT	CTCTTTATGT	AAAGGAAAGT	22080
	TCATGATTGG	ATTTGGCCGG	CCTGACTGCT	TCCCGGCTGT	GATAAAAAAC	ACATGTGAGC	22140
25	TGGGAGGGAA	GTGGGGGAGG	GACACAGCTG	CCCACACAGG	GTTCCACCG	CGGTTACAGG	22200
	GTGGGCAGTG	CTGGGGGAGC	TTTCTCTGTG	GGGGGCTCAG	AGCCTGAGGA	CAGGTGAGCC	22260
	TCTCCGACAC	CTCCCCAGTT	GCCTGGAGTC	TAAACCGTCC	GTTGTCTGTA	CCGTCCGTTT	22320
30	TTCTGCTGA	CTCCTGGTAG	TTCTGAAAG	CTTCTCTTGG	CCAGAGAAGG	GGTTTCAGAG	22380
	GCCGTGTGTC	CAGGCCATT	TGCAAAAGTC	AACTTGACCG	TTCTTTCTCT	TTTCTGGCCT	22440
	GCGTGGTCTG	AAGCTCAGAG	CCCTCTCTTC	ACCCAGCCTG	TGTGTGTCTT	GCCGGACAGA	22500
35	AGAAAAATGG	TGCTTTTTGC	GTGTTAGCAG	AGGTGCTTTT	CATGGCTGAC	CTCAACGCGT	22560
	CCATCTCCAG	CCTTGACCAA	GCTGTTTTTT	AGGGGCAAAC	GCAGGCAAGT	TCTGAATGCA	22620
	CACAGTTATT	TCATGGTTAA	ACTATTGAGC	TTTGGCCGGG	CGCAGTGTGG	CTCTCACGCC	22680
40	TGTCATCCCA	GCACTTTGGG	AGGCCGAGGC	GGGTGGATCA	CCTGAGGTCA	GGAGTTCGAG	22740
	ACCAGCCTGG	CCAACACGGT	GAAACTCTAT	CTCTACTAAA	AATACAAAAA	TTAGCCGGGC	22800
	GTGGTGGTGT	GTATCTGTAA	TCCCAGCTAC	TCAGGAAGCT	GAGGCAGGAG	AATCGCTTGG	22860
45	ACCCAGGAGG	CGGAGGTTGC	ACTGAGCCGA	GATCGCGCCA	TTGCACTCCA	GCCTGGGCGA	22920
	CAGAGCCAGA	CGCTGTCTCA	AAAAAATGAA	TAATAAAATA	AAATAACAGG	AACTAAATAA	22980
	AATAAAACGT	TCAGCTTTGT	TCTGCAAAATC	CACCTCTATT	GTTTACGTTG	GTTTGAGAGA	23040
50	CTCTGTCCCT	TAGAAATAGA	TGTTTGTGTC	CAATTGTAAT	GAATCTGTTT	CAAAAAATGAA	23100
	CAGAATATTC	AAATGGTTTG	AGAGATCTTT	TCCCTTAGAA	ATAGCTTGTT	GCCAATCACA	23160
	AAGAATGTTT	TTCAAAAATG	AATGGAATCT	TCCTGGATAT	CGCTTCCAGA	TCTTCATTTT	23220
	TTTTGCATAG	TTCAACCTGA	AAAGTAAGTG	TCTCAGCCCT	GAATTTCTTT	CTGATTTTTT	23280
55	CATGGGTGT	CTTGCACT	TCTCTGGACT	TGACCACATT	TAAAAAATAA	AAAATTAAC	23340

EP 1 260 228 A2

	TTTTACACAG	GACACGGTTT	CAATAGGAAT	GAGATCTTTG	AGTTTTTATG	TAACAGATTC	23400
5	TTACCATCAG	TTCTCAGATT	CCCAAATTAC	ACACAAAAAG	CCACGGACTT	CGCCTCCTGC	23460
	TAACATGTCC	TTCTGTTTCT	GAGGCTTCTG	TTGGTGTTAG	ACTTTCATGT	TTAATAGCAG	23520
	ACAATGTAGG	GATTTAAAGA	AAAATGCAGA	GAAAGCAAAA	ACACTGACCA	AACACACGGA	23580
10	GATAAGCTTT	CTAAAGCCTT	TGTTCTTGGA	GTTGTCGTTA	AAAAAAAAAA	GTTGTTTTAA	23640
	ACTTTGCAAG	CATGCCCTATA	TTGAACTCAT	AAGCAAGAGA	GCCAAGAAAA	ATAGTGTCGG	23700
	TCGTCTACTC	TACACGTTTT	CCCAAACAG	ACGTATTTTA	ATTTCTTTTG	TTTGAACTCA	23760
	CAGATGCTGA	GAGTTAAAAG	TTAAATTTTT	GTCATGAACA	ATAGTGGCCA	AAACCACAGT	23820
15	TACTTTTGCA	CTATAGCATA	ATAAGAAAAA	TACAGGCTGG	GCTCGGTGGC	TCACACCTGT	23880
	AATCAAAGCA	CTTTTGAGAG	CGAAACAGCC	AGATCCCTTG	AGCCCAGGAG	ATTGAGACCA	23940
	GCCTGGGCAA	CATAGCGAGA	CCCTCATCTC	TACAAAAAAG	GTTTGTTACA	TATGTAACAA	24000
20	ACCTGCACAT	TGTGCACATG	TACCCTAAAA	CTTAAAGTAT	AATAATAAAA	AAATTAAAAA	24060
	AAAATTCACC	AATCAACTGC	CTGCTGGTGC	CTTCAAGAGA	CTCACCTAAC	ACATAAGGAC	24120
	TTGCATAAAC	TTATAAAACA	ATTCAATGGA	AGAATCCTTG	AAAGTATTCT	GAGAAGACAG	24180
25	TATAATAAAC	TGATTCTTAA	AAAGGCTATA	AAAAATTGAA	TAAATCATTG	TTGGGCATCC	24240
	TGTGCTGAAA	TATAATGCAG	CCAATAAAAA	TTACAAAATG	AATAAACATT	TTATAACAAT	24300
	AAAAAAAAGT	CAAATAATTA	GGCAGGCATG	GTGGTGCTCT	CCTACGGTTG	AAGCTATTCA	24360
30	GCAGGCAAGA	GGATACTTTG	TTTTTGTTTT	TTAATTTTTT	TTGAGACAGA	GTCTCGCTCT	24420
	GTTGCCAGGC	TGGAGTGCAG	TGGCGTGATC	TCAGCTCACT	GTAATTTCTG	CCTCCCGGGT	24480
	TCAAGCGATT	TTCTTGCCCC	AGCCTCCCGA	GTAGCTGGGA	TTACAGGTGC	CCGCCACCAC	24540
35	ACCTGGCTAA	TTTCTTTTGT	ATTTTTTAGTA	GAGACGAGGT	TTCCCCATGT	TGGCCAGGCT	24600
	GGTTTTGAGC	TCCCGACCTC	GGGTGATCCA	CCCGCCTCAG	CCTCCCAAAG	TGCTGGGATG	24660
	ACAGGCGTGA	GCCACCGCGC	CTGGCCCAGG	AGGATTATTT	GATCCCAGGA	GGTGGAGGCT	24720
40	GCAGGAAGCC	ATGATTGCAC	CACTGCACTC	CAGCCTGGCT	GACAGAGTGA	GACCACATCT	24780
	CTAAATAAAT	GAATAAATAC	AGGCAGAAAC	TTTTTTTGTT	TTGTTTTGAT	GGAGTCTTGC	24840
	TCTGTCACCA	GGCAGGAGTG	CAGTGGTGCC	ATCTCAGCTC	ACTGCAACCT	CCACCTCCTG	24900
45	GGTTCAAGCA	ATCCTCCTGC	CTCAGCCTCC	CGAGTAGCTG	GGATTACAGG	TGCCC GCCAC	24960
	CACGCCCGGC	TAATTTTTTG	TATGTTTTAGT	AGAGACGGGA	TTTCACCGTG	TTAGCCAGGA	25020
	TGGTCTTGAT	CTCTTGACTT	TGTGATCTGC	CTGCCTCAGC	CTCCCAAAGT	GCTGGGATTA	25080
50	CAGGCATGAG	CCCAGGAGTT	CAAGACCAGC	CTCAGCAACA	AAGTGAGACC	TTTTCTCTCC	25140
	AAAAAATCAA	AAATTTAGCC	AGCTGTGGTG	GCTCCTGCCC	GTGATCCCAG	TACTGTGGGA	25200
	GGCTGAGGCA	GAATTGCTTG	AGCCCAGGAG	TTGAGACCA	ACCTCAGCAA	AAAGGACTCT	25260
55	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTATAT	ATATATATAT	ATATATATAT	25320

EP 1 260 228 A2

	GAGTTTCAAA AATTGCTGGG TGACCAGCTC ATCTACTGGT TTTCCCTTG GGAAAGTGAA	25380
	ATTGTCATGT ATTGAAGATT TCCAAGGAAG TTGTATTGAA TGAGAAACAA ACTCAATCTG	25440
5	TTCGTGTTTA AAGAGCTGCA GTGCGTTTGC TGTGTTTCCC ATAAACTGC ACTTCCAAAA	25500
	GACACGCTGA GAAAGGAGAC CAGGATTTGT AATTCAGAAA TTGGAAAGCA AGTTAGGCTG	25560
	GACGTGGTAG CTCATGCTTG TTGTAATCTC AGCACTCTGG GAGGCTGAGG CAGGAGGATC	25620
10	ACTTGAGCCC AGGAGTTCAA GACCAGCCCG TGCCACATGG TGAAACCCTG TCTCTCCAAA	25680
	AAATAAAACA TTTAGCCAGA TGTGGTGACT CATGCCTGTA ATCCCGGTAT TCTGGGAGGC	25740
	TGAGGCAGAG TTGCTTGAGC CCAGGAGTTC AAGACCAGCC TCGGCAACAA AGTGAGACCC	25800
15	TGTCTCTCCA AAAAAATAAA CATTTAGCCA GCTGTGGTGA CTCATGCCTG TAATCTCAGT	25860
	ACTCTGGGAG GCTGGGGCAG AATGGCTTGA GCCCAGGAGT TCGAGACCAA CCTCAGCAAC	25920
	AAAGTGAGAT CTTGTTTCTC CAAAAATCA AAAATTTAGC CAGCTGTGCT GGCTCATGCC	25980
20	TGTAATCCCG GTACTCTGGG AGGCTGAGGC AGAATCGTTT GAGCCCAGGA GTTCGAGACC	26040
	AACCTCAGCA ACAAAGTGAG ATCTTGTTTC TCCAAAAAAA TCAAAAATTT AGCCAGCTGT	26100
	GCTGGCTGGT GCCTGTAATC CCGGTACTCT GGGAGGCTGA GCGGGAATTG CTTGAGCCCA	26160
25	GGAGTTCAAG ACCAGCCTCA GCAACAAAGT GAGATCTTGT TTCTCCAAA AATAAACAT	26220
	TTAGTCAGCT GTGGTGCTC AAGCCTGTGA TCCCAGCATT TTGGGAGGCC GAGGCGGGCG	26280
	GATCACGAGG TCATGAGATC GAGACCATCC TGGCTAACAC GGTGAAACCC CGTCTCTACT	26340
30	AAAAATACAA AGAAAATTAG CCGGGCGTGG TGGCGGGCGC CTGTAGTCCC AGCTACTCAG	26400
	GAGGCTGAGG CAGGAGAATG CCGTGAGCCT GGGAGGCGGA CCATGCAGTG AGTCAAGATC	26460
	GCGCCACTGC CCTCCAGCCT GGGCCACAGA GCAAGACTCC GTCTCAAAAA AAAAAAAAAA	26520
35	AAAAGTGCTG CCCAACCTGT GTTTGCACCA CTGCCCTCCA GCCTGGGCAA CAGAGCAAGA	26580
	CTCCGTCTCA AAAAAAAAAA AATGCTGCCC AAGCTGTGTT TGCACCACTG CCCTCCAGCC	26640
	TGGGCAACAG AGCAAGACTC CGTCTCAAAA AAAAAAAAAA AAAATGCTGC CCAAGCTGTG	26700
40	TTTGCACCAC TGCCCTCCAG CCTGGGCAAC AGAGCAAGAC TCTGTCTCAA AAAAAAAAAA	26760
	AATGCTGCCC AAGCTGTGTT TGCACCACTG CCCTCCGCC TGGGCAACAG AGCAAGACTC	26820
	CGTCTCAAAA AAAAAAAAAA AATGCTGCCC AAGCTGTGTT TGCACCACTG CCCTCCAGCC	26880
45	TGGGCAACAA AGCAAGCCTC AGCTTTCTGC CATCTCCACA ACCAAGAAAG CAATTACAC	26940
	AGAAATCAGT GCATCGTGCA GTGACCTCTT CAGAAAACCA ATGAGTTTTT CACCTGAGGA	27000
	ACTGTTTCTG AGCCCCATTC AGAAAAACAC ATCCCTGTAA CTGCAGGGCA GATTTACTCA	27060
50	CTGTATGCCT GTTTAAATAA AGCTTCCAGC CTCTGCATGG GGTCTGTCTG GAAGCTCCTG	27120
	TATCTGTCCC ACATTCTTGG AATCACAATG CACCCTTGGG AGGAAGATAT GTATTTAAAG	27180
	GGAGTGATG TTATGGTGAG AAAATGCTGC CCATCCTTCT AGAAGACAAA AGCCACACAA	27240
	AATACATCAC AAGAACCAGT TTTTTCAGA GAAGAACCCTG CACAAAGAAC CTGCTCCCCC	27300
55	CACACCCCA CACACAGGTG AATTAACAGG ATGTATGTTT TATCATAAAA GCACAGGTTT	27360

EP 1 260 228 A2

	GTTTCCTATG CACTCTCTGA GGATTTGGCC ATATGCAAAG ATGTACAAAA ACCTTCTCTT	27420
5	TCCCCAGGGA ACCGTAACCC GTCTGAAAAG ATGCCCTTCT CAGAAGCGAG TTGAACGATT	27480
	GTTGGAAAAG ATAAAATACG ACGTGCACAC ACACAGTAGA GAAATGTCAC CCATGCAAAT	27540
	TATGTGTTTG AATGGAACAC ATTCAGGAAG CTAAATGGGG TATGACCACA CATTTGGGTT	27600
10	GATTTATTTG ACGAGTGGAA GGGGCAGATG GAAATGAATA CTGCTGTTTT CCTTTGGAAG	27660
	GCCATATATG GGAATACCAA GAGGATTACT TTGAAGTTT AGCTTCTCCA GGTGTCTCT	27720
	CTCTCTCTCT CTTTTTTTGA GACAGAGTCT CACTCTGTCA CCCAGGCTGC AGTGCAATGG	27780
15	CGTGCTCTCG GCTCACTGCA ACCTCAGCCT CCCAGGTACA AGCGATTCTC CTGCCCTCAGC	27840
	CTCCCGAGTA GCTGGGATCA CAGGTGTGCA CCACCACGCC TGGCTAATGT TTGTATTTTC	27900
	AGTAGAGATG AGGTTTACC ATGTTGGCCA GGCTGGTCTT GAACTCCTGA CCTCAGGTGA	27960
20	TCCGCCTGCC TCGGCCTCCC AAAGTGCTGG GATGACAGAC ATGAGCTAGC ACGCCCGGCC	28020
	CCAGGTGGTC TTTTGTACGG GTATTAAAGC AGCTTTCTCT CTGAGCCTTA AACCATGAAG	28080
	ATAGACAGAC TCAGTGTATG GGTTTTAGAG TTGTAATTTT ATAAAAATAA GAAAAAGTCG	28140
25	ACCTATCATT GATGGTGTAGT ATTTTTGTGA GCAGTTGCAT GCAATATTAG GATAAGGCAT	28200
	GTTCTCAAAA AGAACTCTTT TTTTTTTTTT TTTGAGACGG AGTCTCGCTC TGTCACCAG	28260
	GCTGGAGTGC AGTGGCACGA TCTCCGCTCA CTGCAAGCTC CTCTTCCCGG GTTCACGCCA	28320
30	TTCTCCTGCC TCAGCCTCCC CAGTAGCTGG GACTACAGGC GCCCGCCACC ACGCCCGGCT	28380
	AATTTTTTGT ATTTTTAGTA GAGACGGGGT TTCACCATGT TAGCCAGGAA GGTCTCGATC	28440
	TCCTGACCTC ATGATCCGTC CGCCTCAGCC TCCCAAAGTG CTGGGACTAC AGGCGTGAGC	28500
35	CACTGCACTT GGCCTTTTTT TTTTTTTAGA TGGAGTTTGT CTCTTGTCGC CCAGGCTGGA	28560
	GTATAATGGC ATGATCTCGA CTCACGTCAA CCTCCGCCTC CCGAGTTCAA GCGATTCTCC	28620
	TGCCTCAGCC TCCCGAGTAG CTGGGATTAC AGGTGCCCAC CACCATGTCA AGATAATGTT	28680
40	TGTATTTTCA GTAGAGATGG GGTGACCA TGTTGGCCAG GCTGGTCTCG AACTCCTGAC	28740
	CTCAGGTGAT CCACCCGCCT TAGCCTCCCA AAGTGCTGGG ATGACAGGCG TGAGCCCCTG	28800
	CGCCCGGCCT TTGTAACCTT ATTTTAAATT TTTTTTTTTT TTTAAGAAAG ACAGAGTCTT	28860
45	GCTCTGTAC CCAGGCTGGA GCACACTGGT GCGATCATAG CTCACTGCAG CCTCAAACCTC	28920
	CTGGGCTCAA GCAATCCTCC CACCTCAGCC TCCTGAGTAG CTGGGACTAC AGGCACCCAC	28980
	CACCACACCC AGCTAATTTT TTTGATTTT ACTAGAGACG GGATCTTGCT TTGCTGCTGA	29040
50	GGCTGGTCTT GAGCTCCTGA GCTCCAAAGA TCCTCTCACC TCCACCTCCC AAAGTGTTAG	29100
	AATTACAAGC ATGAACCACT GCCCGTGGTC TCCAAAAAAA GGAAGTTTAC GTGGATGTTT	29160
	TAGCTTCCTG TTCTCGTCTT TTCTTTGTTA ATGTGTACAG TTGAGGGTGT GTGTGCGTGT	29220
	GCGCACGTGT GTGTGTGCAG TCTCCTGATT TCATGTATTT AATTGTTATT ACCACCACCT	29280
55	CCATCTCTCA TTCCTTCTTA CCCTCACTGT GTAAAGATAC ATGTTGTTTT TAAATTTTAT	29340

EP 1 260 228 A2

	GTATTTATAT TTATTTATTT GTATTTCTGA GACAGAGTCT CACTCTGTG CCCAGGCTAG	29400
	TGGCATGATC TCAGCTCACA GCAACCTTTG CCTCTGGGT TCAAGCGATT CTCCTGCCTC	29460
5	AGCCTCCCGA GTAGCTGAGA TTACAGGCAC ACACCACCAC ACCCGGCTAG TTTTGTTTTG	29520
	AGACGGAGTC TCGCTCTGTT GCAGGCTGCA GTGCAGTGGC GTGATCCTGG CTCACTGCAA	29580
	CCTCTGCCTC CTGGATTCOA GCGATTCTCC TGCCTCAGCC TCCCAAGTAG CTGGGATTAC	29640
10	AGGCGCCAC CGCCACACCT GGCTAATTTT TTATTGGTAG TAGAGACGGG GTTTCTCCAT	29700
	GTTGACCAGA CTGGTCTTGA ACTCCCAACC TCGGGTGATC CACCCACCTG GGCCTCCCAA	29760
	AGTGCTGGGA TGACAGGCGA GGGCCACCGC GTCCAGCCTT CTCTCTCTTC TTCTTTTTTT	29820
15	TTTTTTTAAG ATGGAGTTTC ACTCTGTTGC CCAGGCTGGA GTGCAGTGGT GCAATCTCGG	29880
	CTCCCTGCAA CCTCCACCTC CCAGGTTCAA GAAATTCTTT TGCTCAGCC TCCCGAGTAG	29940
	CTGGGACTAC AGGTGCCCCG CACCACACCC ACCTAATGTT TGTATTTTTT TGGTAGAGAC	30000
20	GGGGCTTCAC CACATTGGCC AGGCTGGTCT TGAACCTCTG ACTTCAGATG ATCCTCCTGC	30060
	CTCAGCCTCC CAGAGTGTG GGATTACAGG CGTGAGCCAC GGTGCCCCGC CAGACGTCAT	30120
	GTCTTAGGAA ATCAGAAAGT GGGTAGTTTC CGCACTCTGA GGAGAAAAAG AGACGTCCGG	30180
25	CGAAGAGAAA GGAGAGTGAA AGGATGTCTC CTCTTGCTCG TAGCCTGTTC TCAATCGTGA	30240
	GTGAGCCAAT TGCCAGAAAC TGAGGGTGCT TCATTGGCC AGGCAAGCTT CTCAACAGAA	30300
	TGTCTAAGTA CTTGTTAATG CTGAGAAGCT CTCCAAGCTA CTGCACTCCA GCCTGGGTGA	30360
30	CAGAGCACGA CCTTGCTCTGA AAACAATTAA TTAATCAATT AATTAATATA ATGAAATCAT	30420
	ACTGAACTCA GGAGACCATT GGGGTGGGCA GGGCTGGGGT TGGAAAGGAA CATAAAATAT	30480
	GGTGCAATGG ACTTTGCTCC AGTCTCCCTC CCCATCTCTT CTCGCCAAGA GTCTCTGGAG	30540
35	GGAGCATGGG GAAGATGCTT TGGGAATCTG TAACTTCTTG TCTTGTAAC AGAATATCTA	30600
	AGTAATTGTT AATGCTGAGA AGTTATAGAT TTCCAAAGCC TTCTCTCAGG CTACGGACAA	30660
	GGGTCATGGG TTACTCAGTG TTACAGAAAG AATGACATGG AGATGTTTGT TACATCTTAA	30720
40	GGAACCATGA GGGGCCAGAG TATTTTACTC TAAGTGTAGA TGGTACATTG GCCACGCCCTG	30780
	TCCCAACACC ACCAATGGTG GCACCTAACT TTGTGTTTG TGCCCCACAT TTCTTCTTCT	30840
	TTTCTGACGT AAATGCAAGT GATATTCCTT GGAAACCATG CTGCAGCAAG AGGCCATCTG	30900
45	ACTACTAGTG ATACCCTGTA GCTCACCTAC AGCAGCTCAC TTGAAGCAGC TCACCCATAG	30960
	CTCAGGTATA GCTCACCTGC AGCGGCTCAC CTGTAGCTCA CGTGTAGCTC ACTTGTAGCA	31020
	GCTCACTGGT AGCTCACCTG CAGCAGCTCA CCTGTACCTC ACCTGTACCT CACCTGCAGC	31080
	AGCTCACCTG TAGCTCACCT GTACGTGAGC CACCGTACCC GGCCAGCAAG ACCCCATTTC	31140
50	TAAAAATAAT ACACAAAAAT TAGCCGGACG CGGTGGCGCG TGTCTGTAGT TGTAGCTACT	31200
	CAGGAGGCTG AGGTGGGAGG ATTGCTGGAG GCTGGGAGGT AGAGGCTGCA GTGAACCGTG	31260
	ATCCAGCCAC TGTACTCTAG CCTGGATGAC ATAGCAAAAC CTGTCTCAA AAAACAAAAA	31320
55	CAAAAAACAA AACAAAGAAA CAAACAAAAA ACCCACACAC ACCGGAAC AAAACAAAAA	31380

EP 1 260 228 A2

5 GCAAAAAGGA AAGAAAAGAG AGCCAGGTCC CAAATATATA TTTCCTTGGA GAACCATTTG 31440
 CAAAGAGCAC ACTTAAGGCC GGGCGCGGTG GCTCACGCCT GTCATCCCGG CACTTTGGGA 31500
 GGCCGAGGTG GGTGGATCAC GAGGTTGGGA GATCGAGACC ATCCTGGCCA ACATGGCGAA 31560
 ACCCCATCTC TACTAAAAAT ACAAAAAATC AGCCAGGTGC TGAGGCAGGT GCCTGTAGTC 31620
 10 CCAGCCACTC AGGAGGCTGA GGCAGGAGAA TGGCATGAAC CTGGGAGGTG GAGGTTGCAG 31680
 TGAGCCGAGA TCGCGCCCCT GCACTCCAGC CTGGGCGACA GAGCGAGACT CCTTCTCAAA 31740
 TAAATAAATA AATAAATAAC AAAGAGCAAA CTTAAATTTG TCTCAGAAAT CCCACGGGAT 31800
 ATTGATCTC CCTCATGCCT ATCTGATGAC ACTTTGAGTG TCTGGGGCCC CGTGCCTATT 31860
 15 TTCTGGGGTT CCCAGAAGCT GCCGTTCTGA AAGTGTGGCT CTCGGGGACG TGGCACAGGT 31920
 GTGGATGTCT GTTTTAAATG TCAGGCGTTT GGACGTTGAG GAACGTGAGG CTGAAGGTCG 31980
 CCTTCGCCGA CCCCTGAGT TTAGGTCCT GCCTTTTAAA ATCTTCCCAG CACTCTGTTG 32040
 20 TTCACGCAAG CGTCCCATCT GTTTGGGTGG CCGTGCCGTC TGCATCTGTC TCGAACCTTC 32100
 ACAGCTTTGC AGAATATCCT GTTTCTCAAT ACGGATGGAG AAACACGAGA CGCGTTTCT 32160
 GGGTTATTTT AGCCGTCACG GAGAACCCCA GACTCATGTG TGCTAATGAC CTCATTAATG 32220
 25 ATACTCTGAG GCAGACAGCC CTGCCTGATC TTAACAACAT TTTTAAATTT TCTTTTTTTG 32280
 TTGTTGTTGT TACAGCATCA TTCATATAAC GTAGGAAACC GTGATCAGTA GCTTTTAGGA 32340
 TATTTGCAAC AGGGTGTAAC ADAAAABD 32367

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 806 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "SHOT"

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 43..615

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

45 GTGTCCCGG AGCTGAAAGA TCGCAAAGAG GATGCGAAAG GGATGGAGGA CGAAGGCCAG 60
 ACCAAAATCA AGCAGAGGCG AAGTCGGACC AATTTCACCC TGAACAACCT CAATGAGCTG 120
 GAGAGGCTTT TTGACGAGAC CCACTATCCC GACGCTTCA TGCAGAGGA ACTGAGCCAG 180
 50 CGACTGGGCC TGTCGGAGGC CCGAGTGCAG GTTTGGTTTC AAAATCGAAG AGCTAAATGT 240
 AGAAAAAAG AAAATCAACT CCATAAAGGT GTTCTCATAG GGGCCGCCAG CCAGTTTGAA 300
 GCTTGTAGAG TCGCACCTTA TGTCACGTA GGTGCTTTAA GGATGCCATT TCAGCAGGTT 360
 55 CAGGCGCAGC TGCAGCTGGA CAGCGCTGTG GCGCACGCGC ACCACCACCT GCATCCGCAC 420

EP 1 260 228 A2

CTGGCCGCGC ACGCGCCCTA CATGATGTTT CCAGCACCGC CCTTCGGACT GCCGCTCGCC 480
 ACGCTGGCCG CGGATTCGGC TTCCGCCGCC TCGGTAGTGG CGGCCGCAGC AGCCGCCAAG 540
 ACCACCAGCA AGGACTCCAG CATCGCCGAT CTCAGACTGA AAGCCAAAAA GCACGCCGCA 600
 GCCCTGGGTC TGTGACVCCA ACGCCAGCAC CAATGTCGCG CCTGTCCCGC GGCACTCAGC 660
 CTGCASNCCC TNDDKANMCG TTRCTYHTCM ATTACACTTT GGGACCYCGG GDBAGVCCTT 720
 TTNNAGACTT YVATKGGSCW CSCTGGBCCC TBRKGAVVAC TTGSGHYCGR GAACCGAKHT 780
 GCCCABAYGA GGACCRGTTT GGAADG 806

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 190 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Glu Asp Glu Gly Gln Thr Lys Ile Lys Gln Arg Arg Ser Arg Thr
 1 5 10 15
 Asn Phe Thr Leu Glu Gln Leu Asn Glu Leu Glu Arg Leu Phe Asp Glu
 20 25 30
 Thr His Tyr Pro Asp Ala Phe Met Arg Glu Glu Leu Ser Gln Arg Leu
 35 40 45
 Gly Leu Ser Glu Ala Arg Val Gln Val Trp Phe Gln Asn Arg Arg Ala
 50 55 60
 Lys Cys Arg Lys Gln Glu Asn Gln Leu His Lys Gly Val Leu Ile Gly
 65 70 75 80
 Ala Ala Ser Gln Phe Glu Ala Cys Arg Val Ala Pro Tyr Val Asn Val
 85 90 95
 Gly Ala Leu Arg Met Pro Phe Gln Gln Val Gln Ala Gln Leu Gln Leu
 100 105 110
 Asp Ser Ala Val Ala His Ala His His His Leu His Pro His Leu Ala
 115 120 125
 Ala His Ala Pro Tyr Met Met Phe Pro Ala Pro Pro Phe Gly Leu Pro
 130 135 140
 Leu Ala Thr Leu Ala Ala Asp Ser Ala Ser Ala Ala Ser Val Val Ala
 145 150 155 160
 Ala Ala Ala Ala Ala Lys Thr Thr Ser Lys Asp Ser Ser Ile Ala Asp
 165 170 175
 Leu Arg Leu Lys Ala Lys Lys His Ala Ala Ala Leu Gly Leu
 180 185 190

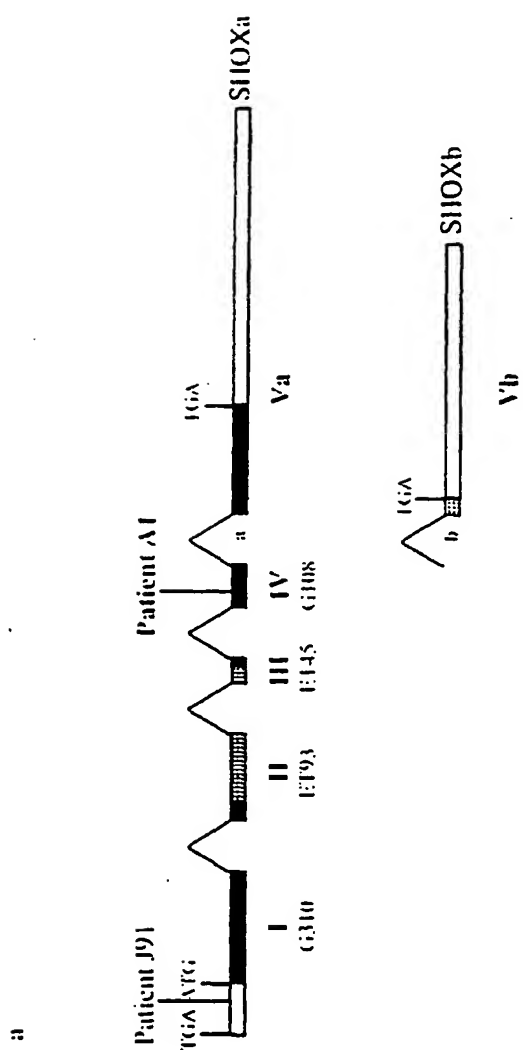
Claims

1. A pharmaceutical composition comprising a protein having regulating activity on human growth, whereby the protein is encoded by a nucleic acid molecule comprising the nucleotide sequence SHOX ET93 [SEQ ID NO: 2] and a nucleotide sequence selected from the group consisting of SHOX G310 [SEQ ID NO: 3], SHOX ET45 [SEQ ID NO: 4], SHOX G108 [SEQ ID NO: 5], SHOX Va [SEQ ID NO: 6] and SHOX Vb [SEQ ID NO: 7] or the nucleotide sequence of SHOT [SEQ ID No.15].
2. A pharmaceutical composition according to claim 1 comprising a protein having the amino acid sequence of SHOXa [SEQ ID NO: 11].
3. A pharmaceutical composition according to claim 1 comprising a protein having the amino acid sequence of SHOXb [SEQ ID NO: 13].
4. A pharmaceutical composition according to claim 1 comprising a protein having the amino acid sequence of SHOT [SEQ ID NO: 16].
5. Use of a protein having regulating activity on human growth, whereby the protein is encoded by a nucleic acid molecule comprising the nucleotide sequence SHOX ET93 [SEQ ID NO: 2] and a nucleotide sequence selected from the group consisting of SHOX G310 [SEQ ID NO: 3], SHOX ET45 [SEQ ID NO: 4], SHOX G108 [SEQ ID NO: 5], SHOX Va [SEQ ID NO: 6] and SHOX Vb [SEQ ID NO: 7], or the nucleotide sequence of SHOT [SEQ ID No. 15] the preparation of a pharmaceutical composition for the treatment of short stature.
6. Use of a protein according to claim 5, the protein having the amino acid sequence of SHOXa [SEQ ID NO: 11].
7. Use of a protein according to claim 5, the protein having the amino acid sequence of SHOXb [SEQ ID NO: 13].
8. Use of a protein according to claim 5, the protein having the amino acid sequence of SHOT [SEQ ID NO: 16].
9. A method for the preparation of a medicament for the *in vivo* treatment of human growth disorders related to a genetic defect in the SHOX or SHOT gene by gene therapy, said SHOX gene having the partial nucleotide sequence as given in [SEQ. ID NO. 8] or having the nucleotide sequence given in [SEQ. ID NO. 14] and said SHOT gene having the nucleotide sequence given in [SEQ. ID NO. 15], the method comprising introducing into an isolated human cell an expression plasmid in which a nucleic acid molecule is incorporated downstream from the expression promoter that effects expression in a human host cell, said nucleic acid molecule comprising the nucleotide sequence SHOX ET93 [SEQ ID NO: 2] and a nucleotide sequence selected from the group consisting of SHOX G310 [SEQ ID NO: 3], SHOX ET45 [SEQ ID NO: 4], SHOX G108 [SEQ ID NO: 5], SHOX Va [SEQ ID NO: 6] and SHOX Vb [SEQ ID NO: 7], or said nucleotide sequence having the sequence SHOT [SEQ ID NO: 15].
10. A method according to claim 9 whereby the nucleic acid molecule encodes a protein having the amino acid sequence of SHOXa [SEQ ID NO: 11].
11. A method according to claim 9 whereby the nucleic acid molecule encodes a protein having the amino acid sequence of SHOXb [SEQ ID NO: 13].
12. A method according to claim 9 whereby the nucleic acid molecule encodes a protein having the amino acid sequence of SHOT [SEQ ID NO: 16].
13. Use of a human growth protein for the preparation of medicaments for the treatment of patients being suspected of having a genetic defect in the human growth gene SHOX, said SHOX gene having the partial nucleotide sequence as given in [SEQ ID NO: 8].
14. Use of a human growth protein for the preparation of medicaments for the treatment of patients being suspected of having a genetic defect in the human growth gene SHOX, said SHOX gene having the nucleotide sequence as given in [SEQ. ID NO. 14].
15. Use of a human growth protein for the preparation of medicaments for the treatment of patients being suspected of having a genetic defect in the SHOT gene, said SHOT gene having the nucleotide sequence as given in [SEQ.

ID NO. 15].

16. Use of a human growth protein according to claims 13 - 15, said patients being identified of having a genetic defect in the human growth gene SHOX or the SHOT gene using a nucleic acid molecule capable of hybridizing to the SHOX or SHOT gene under the following stringent hybridization conditions: 0.5 M NaPi pH 7.2, 7 % SDS and 1 mM EDTA at 65 °C followed by a wash in 40 mM NaPi and 1 % SDS at 65 °C.
17. Use of a human growth protein for the preparation of medicaments for the treatment of short stature in a human subject having a genetic defect in the SHOX or SHOT gene [SEQ. ID NO. 15], said SHOX gene having the partial nucleotide sequence as given in [SEQ. ID NO. 8] or having the nucleotide sequence given in [SEQ. ID NO. 14], said human subject being identified by a method comprising determining said genetic defect in a biological sample isolated from said human subject being susceptible of having a genetic defect in the SHOX or SHOT gene.
18. Use of a human growth protein according to any of claims 13 - 17 wherein the genetic mutation is caused by a hot spot of mutation in the nucleic acid sequence encoding a protein truncation at amino acid position 195 in the SHOX gene, said SHOX gene having the partial nucleotide sequence as given in [SEQ. ID NO. 8] or having the nucleotide sequence given in [SEQ. ID NO. 14].
19. Use of a human growth protein according to claims 13 - 18 wherein the human growth protein is human growth hormone.
20. Use according to claim 19 with the proviso that the preparation of medicaments for the treatment of patients suffering from Turners Syndrome is excluded.

Fig. 1



SHOXa

Fig. 2

```

1   GTGATCCACCCGCGCGCACGGGCCGCTCCTCTCCGCGCGGGGAGACGCGCGCATCCACCAG
61  CCCC GGCTGCTCGCCAGCCCCGGCCCCAGCCATGGAAGAGCTCACGGCTTTTGTATCCAA
      M E E L T A F V S K
121 GTCTTTTGACCAGAAAAGCAAGGACGGTAACGGCGGAGGCGGAGGCGGCGGAGGTAAGAA
      S F D Q K S K D G N G G G G G G G G K K
181 GGATTCCATTACGTACCGGGAAGTTTTGGAGAGCGGACTGGCGCGCTCCCGGGAGCTGGG
      D S I T Y R E V L E S G L A R S R E L G
241 GACGTCGGATTCCAGCCTCCAGGACATCACGGAGGGCGGCGGCCACTGCCCGGTGCATTT
      T S D S S L Q D I T E G G G H C P V H L
301 GTTCAAGGACCACGTAGACAATGACAAGGAGAACTGAAAGAATTCGGCACC GCGAGAGT
      F K D H V D N D K E K L K E F G T A R -V
361 GGCAGAAGGGATTTATGAATGCAAAGAGAAGCGCGAGGACGTGAAGTCGGAGGACGAGGA
      A E G I Y E C K E K R E D V K S E D E D
421 CGGGCAGACCAAGCTGAAATCAGAGGCGCGAGCCGACCAACTTCACGCTGGGAGCAGTGAA
      G - Q T K L K Q - R - R S R T N F T L E Q L N
481 CGAGCTCGAGCGACTTTTTGACGAGACCCATTACCCGACGCCTTCATGCGCGAGGAGCT
      E L E R L F D E T H L Y P D A F M R E E L
541 CAGCCAGCGCCTGGGGCTTTCCGAGGCGCGCGTGCAGGTTTGGTTCCAGAACCGGAGAGC
      S Q R L G L S E A L R V Q V W F Q N R R A
601 CAAGTCCCGCAAACAAGAGATCAGATGCATAAAGGCGTCATCTTGGGCACAGCCAACCA
      K C R K Q E N Q M H K G V I L G T A N H
661 CCTAGACGCCTGCGAGTGGCACCCTACGTCAACATGGGAGCCTTACGGATGCCTTTCCA
      L D A C R V A P Y V N M G A L R M P F Q
721 ACAGGTCCAGGCTCAGCTGCAGCTGGAAGGCGTGGCCCACGCGCACCCGACCTGCACCC
      Q V Q A Q L Q L E G V A H A H P H L P
781 GCACCTGGCGGCGCACGCGCCCTACCTGATGTTCCCCCGCGCCCTTCGGGCTGCCCAT
      H L A A H A P Y L M F P P P P F G L P I
841 CGCGTCGCTGGCCGAGTCCGCCTCGGCCGCGCGCTGGTCGCCGCGCGCCCAAAAGCAA
      A L S A E S A S A A A V V A A A A K S N
901 CAGCAAGAATTCCAGCATCGCCGACCTGCGGCTCAAGGCGCGGAAGCACGCGGAGGCCCT
      S K N S S I A D L R L K A R K H A E A L
961 GGGGCTCTGACCCCGCGCAGCCCCCGCGCGCCCGGACTCCCGGGCTCCGCGCACCCC
      G L *
1021 GCCTGCACCGCGCGCTCCTGCACTCAACCCCGCCTGGAGCTCCTTCCGCGGCCACCGTGCT
1081 CCGGGCACCCCGGAGCTCCTGCAAGAGGCCTGAGGAGGGAGGCTCCCGGGACCGTCCAC
1141 GCACGACCCAGCCAGACCCCTCGCGGAGATGGTGCAGAAGGCGGAGCGGGTGAGCGGCCGT
1201 GCGTCCAGCCCGGGCCTCTCCAAGGCTGCCCGTGCCTCCTGGGACCCCTGGAGAAGGGTAA
1261 ACCCCCGCCTGGCTGCGTCTTCTCTGTATACCTATGCATGCGGTTAACTACACAGT
1321 TTGGAAGATCCTTAGAGTCTATTGAAACTGCAAAGATCCCGAGCTGGTCTCCGATGAAA
1381 ATGCCATTTCTTCGTTGCCAACGATTTTCTTTACTACCATGCTCCTTCCTTCATCCCGAG
1441 AGGCTGCGGAACCGGTGTGATTTGAATGTGGACTTCGGAATCCAGGAGGCAGGGGCCG
1501 GGCTCTCCTCCACCGCTCCCCCGGAGCCTCCAGGCAGCAATAAGGAAATAGTTCTCTGG
1561 CTGAGGCTGAGGAGCTGAACCGCGGGCTTTGGAAAGGGAGGGGAGGAGACCCGAACCTC
1621 CCACGTTGGGACTCCACGTTCCGGGACCTGAATGAGGACCGACTTTATAACTTTTCCA
1681 GTGTTTGATTCCCAAATTGGGTCTGGTTTGTGTTTGGATTGGTATTTTTTTTTTTTTT
1741 TTTTTGCTGTGTACAGGATTACAGACGCAAAAGACTTGATAAGAGACGGACGCGTGGTT
1801 GCAAGGTGTCTACTGATATGCAGCATTAACTTTACTGACATGGAGTGAAGTGCAATATT
1841 ATAAATATTATGATTAATAAAAAAAATAGC [A]

```

SHOXb

Fig. 3

```

1   GTGATCCACCCGCGCGCACGGGCCGTCTCTCCGCGCGGGGAGACGCGCGCATCCACCAG
61  CCCCCGCTGCTCGCCAGCCCCGGCCCCAGCCATGGAAGAGCTCACGGCTTTTGTATCCAA
    M E E L T A F V S K
121 GTCTTTTGACCAGAAAAGCAAGGACGGTAACGGCGGAGGCGGAGGCGGCGGAGGTAAGAA
11  S F D Q K S K D G N G G G G G G G G G K K
181 GGATTCCATTACGTACCGGGAAGTTTTGGAGAGCGGACTGGCGCGCTCCCCGGGAGCTGGG
31  D S I T Y R E V L E S G L A R S R E L G
241 GACGTCGGATTCCAGCCTCCAGGACATCACGGAGGGCGGCGGCCACTGCCCGGTGCATTT
51  T S D S S L Q D I T E G G G H C P V H L
301 GTTCAAGGACCACGTAGACAATGACAAGGAGAACTGAAAGAATTCCGGCACCGCGAGAGT
71  F K D H V D N D K E K L K E F G T A R V
361 GGCAGAAAGGATTATGAATGCAAAGAGAAGCGCGAGGACGTGAAGTCGGAGGACGAGGA
91  A E G I Y E C K E K R E D V K S E D E D
421 CGGGCAGACCAAGCTGAAACAGAGGGCGCAGCCGACCAACTTCACGCTGGAGCAGCTGAA
111 G Q T K L K Q R R S R T N F T L E Q L N
481 CGAGCTCGAGCGACTTTTTGACGAGACCCATTACCCCGACGCCTTCATGCGCGAGGAGCT
131 E L E R L F D E E T H Y P D A F M R E E L
541 CAGCCAGCGCCTGGGGCTTTCCGAGGCGCGCGTGCAGGTTTGGTTCCAGAACC GGAGAGC
151 S Q R L G L S E A R V Q V W F Q N R R A
601 CAAGTCCCGCAAACAAGAGAAATCAGATGCATAAAGGCGTCATCTTGGGCACAGCCAACCA
171 K C R K Q E N Q M H K G V I L G T A N H
661 CCTAGACGCCGTGCCGAGTGGCACCCCTACGTCAACATGGGAGCCTTACGGATGCCTTTCCA
191 L D A C R V A P Y V N M G A L R M P F Q
721 ACAGATGGAGTTTTGCTCTGTGCGCCAGGCTGGAGTATAATGGCATGATCTCGACTCAC
211 Q M E F C S C R P G W S I M A *
781 TGCAACCTCCGCCTCCCGAGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGG
841 ATTACAGGTGCCCACCACCATGTCAAGATAATGTTTGTATTTTCAGTAGAGATGGGGTTT
901 GACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCGCCTTAGCC
961 TCCCAAAGTGCTGGGATGACAGGCGTGAGCCCCGCGCCCGGCCTTTGTAACCTTTATTTT
1021 TAATTTTTTTTCTTTTAAAGAAAGACAGAGTCTTGCTCTGTACCCAGGCTGGAGCACA
1081 CTGGTGGGATCATACCTGCTGAGCCTCAAACCTCCTGGGCTCAAGCAATCCTCCCACCT
1141 CAGCCTCCTGAGTAGCTGGACTACAGGCACCCACCACACCCAGCTAATTTTTTTGA
1201 TTTTACTAGAGACGGGATCTTGCTTTGCTGCTGAGGCTGGTCTTGAGCTCCTGAGCTCC
1261 AAAGATCCTCTCACCTCCACCTCCCAAAGTGTTAGAATTACAAGCATGAACCACTGCCCC
1321 TGGTCTCCAAAAGGACTGTTACGTGG [A].

```

Fig. 4

GTGTCCCGGAGCTGAAAGATCGCAAAGAGGATGCGAAAGGGATGGAGGACGAAGGCCAG
 M E D E G Q
 ACCAAAATCAAGCAGAGGCGAAGTCGGACCAATTTACCCCTGGAACAACTCAATGAGCTG
 T K I K D E R R E R T N E T L E Q L N E L
 GAGAGGCTTTTTGACGAGACCCACTATCCCGACGCCTTCATGCGAGAGGAACTGAGCCAG
 E R L E D E T H Y E D A E M R E E L S Q
 CGACTGGGCCTGTGCGGAGGCCCGAGTGCAGGTTTGGTTTCAAAATCGAAGAGCTAAATGT
 R E G A S E P V Q V W E Q N R R K
 AGAAAACAAGAAATCAACTCCATAAAGGTGTTCTCATAGGGGCCGCCAGCCAGTTTGAA
 R K Q S N Q L H K G V L I G A A S Q F E
 GCTGTAGAGTCGCACCTTATGTCAACGTAGGTGCTTTAAGGATGCCATTCAGCAGGAT
 A C R V A P Y V N V G A L R M P F Q Q D
 AGTCATTGCAACGTGACGCCCTTGCCCTTTCAGGTTTCAGGCGCAGCTGCAGCTGGACAGC
 S H C N V T P L P F Q V Q A Q L Q L D S
 GCTGTGGCGCACGCGCACCACCACCTGCATCCGCACCTGGCCGCGCACGCGCCCTACATG
 A V A H A H H H L H P H L A A H A P Y M
 ATGTTCCAGCACCGCCCTTCGGACTGCCGCTCGCCACGCTGGCCGCGGATTGGGCTTCC
 M F P A P P F G L P L A T L A A D S A S
 GCCGCCTCGGTAGTGGCGGCCGAGCAGCCGCCAAGACCACGCAAGGACTCCAGCATC
 A A S V V A A A A A A K T T S K D S S I
 GCCGATCTCAGACTGAAAGCCAAAAAGCACGCCGAGCCCTGGGTCTGTGACGCCAACGC
 A D L R L K A K K H A A A L G L *
 CAGCACCAATGTGCGGCCTGTCCCGGGCACTCAGCCTGCACGCCCTCCGCGCCCCGCTG
 CTTCTCCGTTACCCCTTTGAGACCTCGGGAGCCGCCCTCTTCCCGCCTCACTGACCATC
 CCTCGTCCCTATCGCATCTTGACTCGGAAAGCCAGACTCCACGCAGGACCAGGGATCT
 CACGAGGCACGACGGCTCCGTGGCTCCTGCCGTTTCTACTCGAGGGCCTAGAATTGG
 GTTTTGTAGGAGCGGGTTTGGGGAGTCTGGAGAGAGACTGGACAGGGTAGTGCTGGAAC
 CGCGGAGTTTGGCTCACCAGCAAGCTACAACGATGGACTCTTGCATAGAAAAAAAATC
 TTGTTAACAATGAAAAAATGAGCAAAACAAAAAATCGAAAGACAAACGGGAGAGAAAAAG
 AGGAAGGCAACTTATTTCTTAAGTCTATTTGGCAGAAGCTGAAATTGGAGAACCAAGGA
 GCAAAAACAAATTTTAAATTAAGTATTTTATACATTTAAAAATATGGAAAAACAACCC
 AGACGATTCTCGAGAGACTGGGGGAGTTACCAACTTAAATGTGTGTTTTAAAAATGCG
 CTAAGAAGGCAAGCAGAAAGAAGAGGTATACTTATTTAAAAAACTAAGATGAAAAAGT
 GCGCAGGTGGGAAGTTACAGGTTTGAAGTACCTTTTCTGCGAAGTTCACGTTAAT
 ACGAGAAATTTGATGAGAGAGGCGGCCCTCCTTTACGTTGAATCAGATGCTTTGAGTTT
 AAACCCACCATGTATGGAAGAGCAAGAAAAAGAGAAAATATTAACGAGGAGAGAGAAAA
 ATAATGGCAAACTGTCTGGACTGCTGACAGTAAATTCGGTTTGCATGGAAAAA
 AAAAAAAAAAAAAAAAAA

Fig. 5

Exon/Intron Organization of the human SHOX gene						
Exon	cDNA-a	cDNA-b	genomic DNA	Exon Size	Intron/Exon	Exon/Intron
I	UTR-368	UTR-368	1-368	368	GTGATCCACC	GTGGCAGAAAGgtaagttcct
II	369-577	369-577	3817-4025	209	ccccacgcagGGATTATGA	GCGCGTGCAGGtaggaaccc
III	578-635	578-635	9851-9908	58	tctccccaagGTTTGGTTCC	ATGCATAAAGgtgggtgtcg
IV	636-724	636-724	10029-10117	89	ttggacacacagGCGTCATCTT	TTTCCAACAGgtagctcact
Va	725-1890	----	13364-14529	1166	gctcccgcaGTCACGGCTC	AAAAAATAGC
Vb	----	725-1349	27154-27778	625	tttttttttagATGGAGTTT	TGTTACGTGG
polyA	> 1891	> 1350	----			

Sizes of exons are given in basepairs; exon sequences are shown in capital letters; donor and acceptor splice sites are underlined. Genomic and cDNA sequences are available via GenBank accession no. X77



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) **EP 1 260 228 A3**

(12)

EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
05.03.2003 Bulletin 2003/10

(51) Int Cl.7: **A61K 38/18**, A61K 48/00,
A61K 38/27

(43) Date of publication A2:
27.11.2002 Bulletin 2002/48

(21) Application number: 02011329.6

(22) Date of filing: 29.09.1997

(84) Designated Contracting States:
**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE**
Designated Extension States:
AL LT LV RO SI

(30) Priority: 01.10.1996 US 27633
16.01.1997 EP 97100583

(62) Document number(s) of the earlier application(s) in
accordance with Art. 76 EPC:
97944906.3 / 0 946 721

(71) Applicant: **Rappold-Hoerbrand, Gudrun**
69118 Heidelberg (DE)

(72) Inventors:
• **Rappold-Hoerbrand, Gudrun, Prof. Dr.**
69118 Heidelberg (DE)
• **Rao, Ercole, Dr.**
64560 Riedstadt (DE)

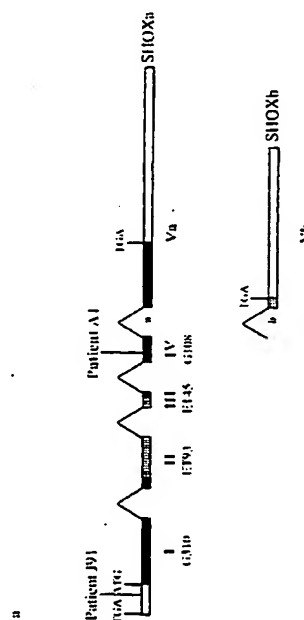
(74) Representative: **Köster, Reinhold, Dr.**
Werderplatz 9
69120 Heidelberg (DE)

(54) Human growth gene and short stature gene region. Therapeutic uses

(57) Subject of the present invention is an isolated human nucleic acid molecule encoding polypeptides containing a homeobox domain of sixty amino acids having the amino acid sequence of SEQ ID NO: 1 and having regulating activity on human growth.

Three novel genes residing within the about 500kb short stature critical region on the X and Y chromosome were identified. At least one of these genes is responsible for the short stature phenotype. The cDNA corresponding to this gene may be used in diagnostic tools, and to further characterize the molecular basis for the short stature-phenotype. In addition, the identification of the gene product of the gene provides new means and methods for the development of superior therapies for short stature.

Fig. 1





European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number

which under Rule 45 of the European Patent Convention EP 02 01 1329 shall be considered, for the purposes of subsequent proceedings, as the European search report

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	THUESTAD INGER JOHANNE ET AL: "Growth hormone treatment in Leri-Weill syndrome." JOURNAL OF PEDIATRIC ENDOCRINOLOGY & METABOLISM, vol. 9, no. 2, March 1996 (1996-03), pages 201-204, XP008011917 * the whole document *	13-20	A61K38/18 A61K48/00 A61K38/27
X	--- DATABASE BIOSIS [Online] BIOSCIENCES INFORMATION SERVICE, PHILADELPHIA, PA, US; 1988 ROSENFELD R.G. ET AL: "Three-year RESULTS OF A randomized Prospective Trial of Methionyl Human Growth Hormone and oxandrolone in turner syndrome" Database accession no. PREV198886106398 XP002225714 * abstract * & JOURNAL OF PEDIATRICS, ISSN: 0022-3476, vol. 113, no. 2, 1998, pages 393-400, --- -/--	13-19	TECHNICAL FIELDS SEARCHED (Int.Cl.6) C07K C12Q C12N A61K
INCOMPLETE SEARCH			
<p>The Search Division considers that the present application, or one or more of its claims, does/do not comply with the EPC to such an extent that a meaningful search into the state of the art cannot be carried out, or can only be carried out partially, for these claims.</p> <p>Claims searched completely :</p> <p>Claims searched incompletely :</p> <p>Claims not searched :</p> <p>Reason for the limitation of the search:</p> <p>Although claims 9-12 (as far as they concern an in vivo method) are directed to a method of treatment of the human/animal body (Article 52(4) EPC), the search has been carried out and based on the alleged effects of the compound/composition.</p>			
Place of search		Date of completion of the search	Examiner
THE HAGUE		20 December 2002	LE CORNEC N.D.R.
CATEGORY OF CITED DOCUMENTS		<p>T : theory or principle underlying the invention</p> <p>E : earlier patent document, but published on, or after the filing date</p> <p>D : document cited in the application</p> <p>L : document cited for other reasons</p> <p>& : member of the same patent family, corresponding document</p>	
<p>X : particularly relevant if taken alone</p> <p>Y : particularly relevant if combined with another document of the same category</p> <p>A : technological background</p> <p>O : non-written disclosure</p> <p>P : intermediate document</p>			

EPO FORM 1503 03 82 (P04C07)



European Patent
Office

Application Number
EP 02 01 1329

CLAIMS INCURRING FEES

The present European patent application comprised at the time of filing more than ten claims.

- ☐ Only part of the claims have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims and for those claims for which claims fees have been paid, namely claim(s):
- ☐ No claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims.

LACK OF UNITY OF INVENTION

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

see sheet B

- ☒ All further search fees have been paid within the fixed time limit. The present European search report has been drawn up for all claims.
- ☐ As all searchable claims could be searched without effort justifying an additional fee, the Search Division did not invite payment of any additional fee.
- ☐ Only part of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the inventions in respect of which search fees have been paid, namely claims:
- ☐ None of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims, namely claims:



European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number
EP 02 01 1329

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
D,A	E. RAO ET AL: "Construction of a cosmid contig spanning the short stature candidate region in the pseudoautosomal region PAR 1." TURNER SYNDROME IN A LIFE SPAN PERSPECTIVE: RESEARCH AND CLINICAL ASPECTS. PROCEEDINGS OF THE 4TH INTERNATIONAL SYMPOSIUM ON TURNER SYNDROME, GOTHENBURG, SWEDEN,, 18 - 21 May 1995, pages 19-24, XP002052955 EDITED BY ALBERTSO-WIKLAND K, RANKE MB * the whole document *	1-18	
A	M. MARRA ET AL: "mj75d03.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 481925 5' similar to TR:G1002494 G1002494 ARIX1." EMBL DATABASE ENTRY MMA59929, ACCESSION NUMBER AA059929, 24 September 1996 (1996-09-24), XP002052953		TECHNICAL FIELDS SEARCHED (Int.Cl.6)
A	M. MARRA ET AL: "mb68b03.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 334541 5'similar to SW: HPR1-chick q05437 homeobox protein PRX-1." EMBL DATABASE ENTRY MM3349, ACCESSION NUMBER W1818334, 4 May 1996 (1996-05-04), XP002052954		
A	L. HILLIER ET AL: "zb81a08.s1 Homo sapiens cDNA clone 309974 3' similar to PIR:S29087 S29087 homeotic protein Otx1-mouse" EMBL DATABASE ENTRY HS100314, ACCESSION NUMBER N99100, 19 April 1996 (1996-04-19), XP002052956 --- -/--		



European Patent
Office

LACK OF UNITY OF INVENTION
SHEET B

Application Number

EP 02 01 1329

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

1. Claims: 1-12 totally and 13-18 partially

Use of SHOXa, SHOXb or SHOT and nucleic acid encoding them in the preparation of pharmaceutical compositions. Gene therapy.

2. Claims: 13-18 partially and 19-20 totally

Use of a human growth hormone for the preparation of medicaments for the treatment of patients having a genetic defect in the SHOX or SHOT gene.



European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number

EP 02 01 1329

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
D,A	T. OGATA ET AL: "short stature in a girl with partial monosomy of the pseudoautosomal region distal to DXYS15: further evidence for the assignment of the critical region for a pseudoautosomal growth gene(s)." JOURNAL OF MEDICAL GENETICS, vol. 32, no. 10, October 1995 (1995-10), pages 831-834, XP002052957 ---		TECHNICAL FIELDS SEARCHED (Int.Cl.6)
A	DATABASE EMBL [Online] 7 September 1996 (1996-09-07) A.C. ROSCAVELLI ET AL: "Mus musculus OG12b homeodomain protein (OG-12) mRNA, complete cds." retrieved from EBI, HINXTON, UK Database accession no. U67055 XP002215676 * abstract * & A.C. ROSCAVELLI ET AL: "Cloning and characterization of four murine homeobox genes" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA., vol. 93, no. 20, 1 October 1996 (1996-10-01), pages 10691-10696, XP002052968 NATIONAL ACADEMY OF SCIENCE. WASHINGTON., US ISSN: 0027-8424 ---		
D,A	A. HENKE ET AL: "Deletions within the pseudoautosomal region help map three new markers and indicate a possible role of this region in linear growth" AMERICAN JOURNAL OF HUMAN GENETICS, vol. 49, no. 4, October 1991 (1991-10), pages 811-819, XP002052958 ---		
	-/--		



European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number
EP 02 01 1329

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
A	B.W. SCHÄFER ET AL: "Molecular cloning and characterization of a human PAX-7 cDNA expressed in normal and neoplastic myocytes." NUCLEIC ACIDS RESEARCH., vol. 22, no. 22, 1994, pages 4574-4582, XP002052959 OXFORD GB		
P,D, X	--- E. RAO ET AL: "Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome." NATURE GENETICS, vol. 16, no. 1, April 1997 (1997-04), pages 54-63, XP002052960 * the whole document *	1-18	TECHNICAL FIELDS SEARCHED (Int.Cl.6)
P,X	--- J.W. ELLISON ET AL: "PHOG, a candidate gene for involvement in the short stature of Turner Syndrome." HUMAN MOLECULAR GENETICS, vol. 6, no. 8, August 1997 (1997-08), pages 1341-1347, XP002052961 * the whole document *	1-18	
P,A	--- A.C. ROVESCALLI ET AL: "Cloning and characterization of four murine homeobox genes" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA., vol. 93, 1 October 1996 (1996-10-01), pages 10691-10696, XP002052968 WASHINGTON US * figures 3,4,6 *	1-18	
	--- -/--		



European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number

EP 02 01 1329

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
T	RAPPOLD G: "SHOX MUTATIONS CAUSE GROWTH FAILURE IN TURNER AND LERI-WEILL SYNDROME" HORMONE RESEARCH, S. KARGER AG, BASEL, CH, vol. 51, no. SUPPL 2, 1999, page 6 XP001002410 ISSN: 0301-0163 ---		
T	VUGUIN P ET AL: "The effect of growth hormone treatment in idiopathic short stature with SHOX mutation" PEDIATRIC RESEARCH, WILLIAMS AND WILKINS, BALTIMORE, MD., US, vol. 43, no. 4 PART 2, April 1998 (1998-04), page 87A XP002168567 ISSN: 0031-3998 ---		TECHNICAL FIELDS SEARCHED (Int.Cl.6)
T	SCHWARZE C P ET AL: "SHOX GENE MUTATIONS IN CHILDREN WITH IDIOPATHIC SHORT STATURE - SCREENING AND THERAPY WITH RHGH" HORMONE RESEARCH, S. KARGER AG, BASEL, CH, vol. 51, no. SUPPL 2, 1999, page 34 XP001002411 ISSN: 0301-0163 -----		

This Page is inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images
problems checked, please do not report the
problems to the IFW Image Problem Mailbox**